

Randomized Matrix Computations *

Victor Y. Pan^{[1,2],[a]}, Guoliang Qian^{[2],[b]}, and Ai-Long Zheng^{[2],[c]}

^[1] Department of Mathematics and Computer Science
Lehman College of the City University of New York
Bronx, NY 10468 USA

^[2] Ph.D. Programs in Mathematics and Computer Science
The Graduate Center of the City University of New York
New York, NY 10036 USA

^[a] victor.pan@lehman.cuny.edu

<http://comet.lehman.cuny.edu/vpan/>

^[b] gqian@gc.cuny.edu

^[c] azheng-1999@yahoo.com

Abstract

Random matrices tend to be well conditioned, and we employ this well known property to advance matrix computations. We prove that our algorithms employing Gaussian random matrices are efficient, but in our tests the algorithms have consistently remained as powerful where we used sparse and structured random matrices, defined by much fewer random parameters. We numerically stabilize Gaussian elimination with no pivoting as well as block Gaussian elimination, precondition an ill conditioned linear system of equations, compute numerical rank of a matrix without pivoting and orthogonalization, approximate the singular spaces of an ill conditioned matrix associated with its largest and smallest singular values, and approximate this matrix with low-rank matrices, with applications to its 2×2 block triangulation and to tensor decomposition. Some of our results and techniques can be of independent interest, e.g., our estimates for the condition numbers of random Toeplitz and circulant matrices and our variations of the Sherman–Morrison–Woodbury formula.

2000 Math. Subject Classification: 15A52, 15A12, 15A06, 65F22, 65F05, 65F10

Key Words: Random matrices, Preconditioning, Numerical rank

1 Introduction

It is well known that random matrices tend to be well conditioned [D88], [E88], [ES05], [CD05], [SST06], [B11], and we employ this property to advance matrix computations. We prove that with probability 1 or near 1 our techniques of randomized preprocessing precondition a large and important class of ill conditioned matrices. By employing randomization we stabilize numerically Gaussian elimination with no pivoting as well as block Gaussian elimination, precondition an ill conditioned linear system of equations, compute numerical rank of a matrix using no pivoting and

*Some results of this paper have been presented at the ACM-SIGSAM International Symposium on Symbolic and Algebraic Computation (ISSAC '2011), San Jose, CA, 2011, the 3rd International Conference on Matrix Methods in Mathematics and Applications (MMMA 2011) in Moscow, Russia, June 22-25, 2011, the 7th International Congress on Industrial and Applied Mathematics (ICIAM 2011), in Vancouver, British Columbia, Canada, July 18-22, 2011, the SIAM International Conference on Linear Algebra, in Valencia, Spain, June 18-22, 2012, and the Conference on Structured Linear and Multilinear Algebra Problems (SLA2012), in Leuven, Belgium, September 10-14, 2012

orthogonalization, approximate the singular spaces of an ill conditioned matrix A associated with its largest and smallest singular values, approximate this matrix by low-rank matrices, compute its 2×2 block triangulation, and compute a Tensor Train approximation of a tensor. Our analysis and experiments show substantial progress versus the known algorithms. In our tests most of our techniques have fully preserved their power when we dramatically decreased the number of random parameters involved, which should motivate further research. Some of our results and techniques can be of independent interest, e.g., our estimates for the condition numbers of random Toeplitz and circulant matrices and our extensions of the Sherman–Morrison–Woodbury formula.

1.1 Numerically safe Gaussian elimination with no pivoting

Hereafter “flop” stands for “arithmetic operation”, “expected” and “likely” mean “with probability 1 or close to 1”, $\sigma_j(A)$ denotes the j th largest singular value of an $n \times n$ matrix A , and the ratio $\kappa(A) = \sigma_1(A)/\sigma_\rho(A)$ for $\rho = \text{rank}(A)$ denotes its condition number. $\kappa(A) = \|A\| \|A^{-1}\|$ if $\rho = n$, that is if A is a nonsingular matrix. If this number is large in context, then the matrix A is *ill conditioned*, otherwise *well conditioned*. For matrix inversion and solving linear systems of equations the condition number represents the output magnification of input errors,

$$\kappa(A) \approx \frac{\|\text{OUTPUT ERROR}\|}{\|\text{INPUT ERROR}\|}, \quad (1.1)$$

and backward error analysis implies similar magnification of rounding errors [GL96], [H02], [S98].

To avoid dealing with singular or ill conditioned matrices in Gaussian elimination, one incorporates pivoting, that is row or column interchange. *Gaussian elimination with no pivoting* (hereafter we refer to it as *GENP*) can easily fail in numerical computations with rounding errors, except for some special input classes such as the classes of diagonally dominant and positive definite matrices. For such matrices, GENP outperforms Gaussian elimination with pivoting [GL96, page 119]. By extending our previous study in [PGMQ, Section 12.2] and [PQZa], we expand these classes dramatically by proving that pre- and post-multiplication of a well conditioned coefficient matrix of full rank by a square Gaussian random matrix is expected to yield safe numerical performance of GENP as well as block Gaussian elimination (see Remark 4.1). The results of our tests support this theoretical finding consistently. Furthermore the tests show that the power of our preprocessing is fully preserved where we use just circulant or Householder multipliers generated by a vector or a pair of vectors, respectively, and where we fill these vectors with integers ± 1 and limit randomization to the choice of the signs \pm (see our Table 10.6 and [PQZa, Table 2]).

1.2 Randomized preconditioning: the basic theorem

Given an ill conditioned matrix A , can we extend our progress by applying randomized multipliers X and Y to yield a much better conditioned matrix product XAY ? No, because random square matrices X and Y are expected to be nonsingular and well conditioned [D88], [E88], [ES05], [CD05], [SST06], [B11] and because $\kappa(XAY) \geq \frac{\kappa(A)}{\kappa(X)\kappa(Y)}$. Approximate inverses are popular multipliers but only where it is not costly to compute them, that is only for some special, although important classes of matrices A , except for the surprising empirical technique in [R90] (see our Remark 9.6).

We can readily produce a well conditioned matrix C by applying additive preprocessing $A \Rightarrow C = A + P$, e.g., by choosing $P = I - A$, but it is not clear how this could help us to solve a linear system $A\mathbf{y} = \mathbf{b}$. (Here and hereafter I denotes the identity matrix.) Assume, however, that we are given a nonsingular ill conditioned $n \times n$ matrix A together with a small upper bound r on its numerical nullity $\text{nnul}(A)$, that is on the number of its singular values that are much smaller than the 2-norm $\|A\|_2$. Such matrices make up a large and important subclass of nonsingular ill conditioned matrices (cf. [CDG03] and Remark 2.1), for which randomized additive preconditioning is supported by the following theorem. In Section 5 we prove it, extend it to rectangular matrices A , and further detail its estimates.

Theorem 1.1. *Suppose A is a real $n \times n$ matrix having a numerical rank ρ (that is the ratio $\sigma_{\rho+1}(A)/\|A\|$ is small, but the ratio $\sigma_\rho(A)/\|A\|$ is not small), σU and σV are standard Gaussian random $n \times r$ matrices, whose all $2nr$ entries are independent of each other, $0 < \rho < n$, $0 < r < n$, and $C = A + UV^T$. Then (i) we can expect that*

$$0.5\|A\|_2 \leq \|C\|_2 \leq 1.5\|A\|_2 \text{ if, say } \sigma/\|A\|_2 \leq \frac{1}{10}. \quad (1.2)$$

(ii) Furthermore the matrix C is singular or ill conditioned if $r < n - \rho$, (iii) but otherwise it is nonsingular with probability 1 and (iv) is expected to be well conditioned if the ratio $\sigma/\|A\|_2$ is neither large nor small, e.g., if $\frac{1}{100} \leq \sigma/\|A\|_2 \leq 100$.

1.3 Randomized algorithms

We recall some known randomized matrix algorithms [D83], [HMT11] and study some new ones. Suppose we are given a normalized nonsingular $n \times n$ matrix A , such that $\|A\| = 1$, and its small positive numerical nullity $r = \text{nnul}(A)$ (cf. Section 8.2 on computing numerical nullity). Suppose also that we have generated two standard Gaussian random $n \times r$ matrices U and V and applied *randomized additive preconditioning* $A \Rightarrow C = A + UV^T$ producing nonsingular and well conditioned matrix C , as we can expect by virtue of Theorem 1.1. Then we can apply the Sherman–Morrison–Woodbury formula, hereafter referred to as the *SMW formula*, to reduce an ill conditioned linear system $A\mathbf{y} = \mathbf{b}$ of n equations to well conditioned linear systems $C\mathbf{x} = \mathbf{f}$.

By virtue of (1.1) we must perform our computations with high accuracy to obtain meaningful output where the matrix A is ill conditioned, but we can limit highly accurate computations with multiple or infinite precision to $O(n^2)$ flops and performs the order of n^3 remaining flops with the standard double or single IEEE precision versus order of n^3 high precision flops in Gaussian elimination. It may be even more promising to combine randomized additive preprocessing and the SMW formula to compute multiplicative preconditioners of the matrix A (see Remark 9.6).

We also apply our approach numerically, with double precision, to the approximation of the singular spaces associated with the $\rho = n - r$ largest and the r smallest singular values of an ill conditioned matrix A that has a positive numerical nullity $r = \text{nnul}(A)$. We rely on the following observations. Suppose the matrix C of Theorem 1.1 is nonsingular and well conditioned, as expected. Then (a) we can readily compute the $n \times r$ matrices $C^{-1}U$ and $C^{-T}V$ by solving $2r$ linear systems of equations with the matrices C and C^T and (b) the ranges of the matrices $C^{-1}U$ and $C^{-T}V$ approximate closely the left and right trailing singular spaces, respectively, associated with the r smallest singular values of an ill conditioned matrix A . Furthermore we approximate the left and right leading singular spaces associated with the ρ largest singular values of the matrix A , both by extending these techniques and by means of random sampling $A \Rightarrow AH$ and $A \Rightarrow A^T G$ for Gaussian random $n \times \rho_+$ matrices G and H and for small positive integers $\rho_+ - \rho$ [HMT11]. We extend these randomized algorithms to compute numerical rank of a matrix using no pivoting or orthogonalization, its 2×2 block triangulation and its low-rank approximation with a further extension to tensor decomposition.

1.4 Sparse and structured matrix computations, randomized augmentation, and numerical experiments

We cannot extend our proof of Theorem 1.1 to the case of Gaussian random Toeplitz matrices U and V , but such extension has been supported consistently by the results of our numerical tests. In these tests our algorithms remained as efficient where we replaced Gaussian random matrices by matrices defined by much fewer random parameters such as the circulant and Householder multipliers of Section 1.1 and the block vectors U with blocks $\pm I$ and O in the additive preprocessing $A \Rightarrow C = A + UV^T$ in Section 10.2. In these cases we limited randomization to the choice of the block sizes in the block vectors and of the signs \pm . We have amended our algorithms to preserve matrix sparseness and structure, in particular for computing numerical rank and nullity.

Additive preprocessing $A \Rightarrow C = A + UV^T$ preserves matrix structure and sparseness quite well where the matrices U and V have consistent structure and sparseness, but both random sampling and randomized augmentation, such as the maps $A \Rightarrow A^T G$ and $A \Rightarrow K = \begin{pmatrix} W & V^T \\ -U & A \end{pmatrix}$, achieve this even more perfectly where we properly extend the patterns of sparseness and structure of the matrix A to the random matrices G , U , V , and W . For $V = -U$ and symmetric positive definite matrices A and K the above augmentation increases the condition number $\kappa(A)$, in contrast to additive preprocessing $A \Rightarrow C = A + VV^T$ where A is a nonnegative definite matrix. Other than that, however, we observe close similarity between augmentation and additive preprocessing, we prove similar preconditioning properties of both maps under randomization and extend to augmentation the SMW and dual SMW formulae as well as Theorem 1.1 and our expressions for the bases of leading and trailing singular spaces (see Section 6, Corollaries 7.2 and 7.4, and equations (6.2) and (7.4)). Then again our tests were in good accordance with all these extensions even in the case of sparse and structured matrices G , U , V , and W , defined by a small number of random parameters, although our formal proofs only apply to the case of Gaussian random matrices G , U , V , and W .

Other than that our tests for randomized multiplicative and additive preprocessing and augmentation were in good accordance with our theoretical estimates. By applying randomized additive preprocessing to ill conditioned matrices we consistently observed dramatic decrease of the condition numbers (see Table 10.5), and this enabled accurate solution of the respective ill conditioned linear systems of equations (see Tables 10.7–10.10 and 10.15), whereas random circulant multipliers filled with ± 1 have stabilized GENP numerically (see Table 10.6). Furthermore according to our test results, our algorithms approximated accurately the leading and trailing singular spaces of ill conditioned matrices and approximated a matrix by a low-rank matrix (see Tables 10.11–10.14). We have also matched the output accuracy of the customary algorithms for solving ill conditioned Toeplitz linear systems of equations but outperformed them dramatically in terms of the CPU time (see Table 10.16). Finally our experimental data on the condition numbers of Gaussian random Toeplitz and circulant matrices are in good accordance with our formal estimates (see Sections 3.4 and 3.5 and Tables 10.1–10.4). These estimates respond to a challenge in [SST06] and can be surprising because the paper [BG05] has proved that the condition numbers of $n \times n$ Toeplitz matrices grow exponentially in n as $n \rightarrow \infty$ for some large and important classes of Toeplitz matrices, whereas we prove the opposit for Gaussian random Toeplitz matrices.

1.5 Organization of the paper and selective reading

We recall some definitions and basic results on matrix computations in the next section. We estimate the condition numbers of Gaussian random general, Toeplitz and circulant matrices in Section 3 and of randomized matrix products in Section 4. The latter estimate imply that randomized multipliers support GENP and block Gaussian elimination. In Section 5 we prove that our randomized additive preprocessing of an ill conditioned matrix is expected to produce a well conditioned matrix. In Section 6 we prove a similar property of randomized augmentation, which we link to randomized additive preprocessing and apply to the solution of ill conditioned Toeplitz linear systems of equations.

In Sections 7–9 we apply randomization to computations with ill conditioned matrices having small numerical nullities or ranks. We compute numerical rank of such matrix using no pivoting or orthogonalization, approximate its trailing and leading singular spaces, approximate it by a low-rank matrix, and point out applications to tensor decomposition and to approximation by structured matrices. We also apply our randomized additive preconditioning to compute 2×2 block triangulation of an ill conditioned matrix A , to compute its inverse, and to precondition a linear system $A\mathbf{y} = \mathbf{b}$. We comment on randomized computations with structured and sparse inputs in Section 8.

In Section 10 we cover numerical tests, which constitute the contribution of the second and the third authors. In Section 11 we comment on the related works, our progress, and some directions for further study. In Appendix A we recall some results on structured matrices. In Appendix B we estimate the probability that a random matrix has full rank under the uniform probability distribution. In Appendix C we comment on the extension of our probabilistic estimates to the case

of complex matrices.

The paper can be partitioned into two parts: the next five sections, Section 11 and the Appendix cover basic theorems and make up Part 1, whereas Sections 7–10, on the algorithms and tests, make up Part 2. The correctness proofs of the algorithms of Part 2 employ the results of Part 1, but otherwise Part 2 can be read independently of Part 1. In *selective reading* one can skip the subjects of structured matrices and tensors (in particular Sections 2.7, 2.8, 3.4, 3.5, 6.5, 8, 10.3, 10.6, and Appendix A) or augmentation (Section 6 and all related materials), or can read only selected algorithmic material, e.g., on low-rank approximation by means of random sampling (Proto-Algorithms 7.1 and 8.1, Section 10.4 and the supporting results) or on preprocessing that supports GENP and block Gaussian elimination (Theorems 2.5, 3.2 and 4.1, Remark 4.1 and Section 10.3). In the paper, unlike its introduction, we study the general case of rectangular input matrices A , but again one may restrict oneself to the simpler special case of square matrices.

2 Some definitions and basic results

We assume computations in the field \mathbb{R} of real numbers, and comment on the extension to the field \mathbb{C} of complex numbers in Appendix C.

Hereafter “flop” stands for “arithmetic operation”; “expected” and “likely” mean “with probability 1 or close to 1”, and the concepts “large”, “small”, “near”, “closely approximate”, “ill conditioned” and “well conditioned” are quantified in the context. For two scalars a and b we write $a \ll b$ and $b \gg a$ if the ratio $|b/a|$ is large. We write $a \approx b$ if $|a - b| \ll |a| + |b|$. Next we recall and extend some customary definitions of matrix computations [GL96], [S98].

2.1 Some basic definitions on matrix computations

$\mathbb{R}^{m \times n}$ is the class of real $m \times n$ matrices $A = (a_{i,j})_{i,j}^{m,n}$.

$(B_1 \mid \dots \mid B_k) = (B_j)_{j=1}^k$ is a $1 \times k$ block matrix with blocks B_1, \dots, B_k . $\text{diag}(B_1, \dots, B_k) = \text{diag}(B_j)_{j=1}^k$ is a $k \times k$ block diagonal matrix with diagonal blocks B_1, \dots, B_k .

\mathbf{e}_i is the i th coordinate vector of dimension n for $i = 1, \dots, n$. These vectors define the identity matrix $I_n = (\mathbf{e}_1 \mid \dots \mid \mathbf{e}_n)$ and the reflection matrix $J_n = (\mathbf{e}_n \mid \dots \mid \mathbf{e}_1)$, both of size $n \times n$. $O_{k,l}$ is the $k \times l$ matrix filled with zeros. $\mathbf{0}_k$ is the vector $O_{k,1}$. We write I , J , O , and $\mathbf{0}$ where the size of a matrix or a vector is not important or is defined by context. Furthermore we write

$$I_{g,h} = I_g \text{ where } g \leq h, \text{ whereas } I_{g,h} = (I_h \mid O_{h,g-h}) \text{ where } g > h. \quad (2.1)$$

A^T is the transpose of a matrix A . A^H is its Hermitian transpose. A matrix A is symmetric if $A = A^T$ and is symmetric positive definite if $A = B^T B$ for a real nonsingular matrix B .

A real matrix Q is called *orthogonal* if $Q^T Q = I$ or $Q Q^T = I$. More generally, over the complex field \mathbb{C} a matrix U is called *unitary* if $U^H U = I$ or $U U^H = I$. Hereafter $Q(A)$ denote a unique orthogonal matrix specified by the following result.

Fact 2.1. [GL96, Theorem 5.2.2]. *QR factorization* $A = QR$ of a matrix A having full column rank into the product of an orthogonal matrix $Q = Q(A)$ and an upper triangular matrix $R = R(A)$ is unique provided that the factor R is a square matrix with positive diagonal entries.

2.2 Range, null space, rank, nullity, nmbs, and generic rank profile

$\mathcal{R}(A)$ denotes the range of an $m \times n$ matrix A , that is the linear space $\{\mathbf{z} : \mathbf{z} = A\mathbf{x}\}$ generated by its columns. $\mathcal{N}(A)$ denotes its null space $\{\mathbf{v} : A\mathbf{v} = \mathbf{0}\}$, $\text{rank}(A) = \dim \mathcal{R}(A)$ its rank, and $\text{nul}(A) = \dim \mathcal{N}(A) = n - \text{rank}(A)$ its right nullity or just *nullity*, whereas $\text{nul}(A^T) = m - \text{rank}(A)$ is the left nullity of A , equal to $\text{nul}(A)$ if and only if $m = n$. \mathbf{v} is the null vector of A if $A\mathbf{v} = \mathbf{0}$.

Fact 2.2. *The set \mathbb{M} of $m \times n$ matrices of rank ρ is an algebraic variety of dimension $(m + n - \rho)\rho$.*

Proof. Let M be an $m \times n$ matrix of a rank ρ with a nonsingular leading $\rho \times \rho$ block M_{00} and write $M = \begin{pmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{pmatrix}$. Then the $(m - \rho) \times (n - \rho)$ Schur complement $M_{11} - M_{10}M_{00}^{-1}M_{01}$ must vanish, which imposes $(m - \rho)(n - \rho)$ algebraic equations on the entries of M . Similar argument can be applied where any $\rho \times \rho$ submatrix of the matrix M (among $\binom{m}{\rho} \binom{n}{\rho}$ such submatrices) is nonsingular. Therefore $\dim \mathbb{M} = mn - (m - \rho)(n - \rho) = (m + n - \rho)\rho$. \square

A matrix B is a *matrix cover* for its range $\mathcal{R}(B)$. A matrix cover is a *matrix basis* (for its range) if it has full column rank. A matrix basis B for the null space $\mathcal{N}(A)$ is a *null matrix basis* or a *nmb* for the matrix A , and we write $B = \text{nmb}(A)$. $\mathcal{N}(A^T)$ is the left null space of a matrix A , and similarly the map $A \Rightarrow A^T$ defines left null vectors, left nmbs, and the left nullity of a matrix A . $A_k^{(k)}$ denotes the leading, that is northwestern $k \times k$ block submatrix of a matrix A . A matrix of a rank ρ has *generic rank profile* if all its leading $i \times i$ blocks are nonsingular for $i = 1, \dots, \rho$. If such matrix is nonsingular itself, then it is called *strongly nonsingular*.

2.3 Norms, SVD, and singular spaces

$\|A\|_h$ is the h -norm and $\|A\|_F = \sqrt{\sum_{i,j=1}^{m,n} |a_{i,j}|^2}$ is the Frobenius norm of a matrix $A = (a_{i,j})_{i,j=1}^{m,n}$. We write $\|A\| = \|A\|_2$ and $\|\mathbf{v}\| = \sqrt{\mathbf{v}^T \mathbf{v}} = \|\mathbf{v}\|_2$ and recall from [GL96, Section 2.3.2 and Corollary 2.3.2] that

$$\max_{i,j=1}^{m,n} |a_{i,j}| \leq \|A\| = \|A^T\| \leq \sqrt{mn} \max_{i,j=1}^{m,n} |a_{i,j}|,$$

$$\frac{1}{\sqrt{m}} \|A\|_1 \leq \|A\| \leq \sqrt{n} \|A\|_1, \quad \|A\|_1 = \|A^T\|_\infty, \quad \|A\|^2 \leq \|A\|_1 \|A\|_\infty, \quad (2.2)$$

$$\|A\| \leq \|A\|_F \leq \sqrt{n} \|A\|, \quad (2.3)$$

$$\|AB\|_h \leq \|A\|_h \|B\|_h \text{ for } h = 1, 2, \infty \text{ and any matrix product } AB. \quad (2.4)$$

A matrix A is *normalized* if $\|A\| = 1$. A normalized vector is orthogonal (unitary), and we call it *unit*. We write $A \approx B$ if $\|A - B\| \ll \|A\| + \|B\|$.

Define an *SVD* or *full SVD* of an $m \times n$ matrix A of a rank ρ as follows,

$$A = S_A \Sigma_A T_A^T. \quad (2.5)$$

Here $S_A S_A^T = S_A^T S_A = I_m$, $T_A T_A^T = T_A^T T_A = I_n$, $\Sigma_A = \text{diag}(\hat{\Sigma}_A, O_{m-\rho, n-\rho})$, $\hat{\Sigma}_A = \text{diag}(\sigma_j(A))_{j=1}^\rho$, $\sigma_j = \sigma_j(A) = \sigma_j(A^T)$ is the j th largest singular value of a matrix A for $j = 1, \dots, \rho$, and we write $\sigma_j = 0$ for $j > \rho$. These values have the minimax property

$$\sigma_j = \max_{\dim(\mathbb{S})=j} \min_{\mathbf{x} \in \mathbb{S}, \|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|, \quad j = 1, \dots, \rho, \quad (2.6)$$

where \mathbb{S} denotes linear spaces [GL96, Theorem 8.6.1]. Consequently $\sigma_\rho > 0$, $\sigma_1 = \max_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\| = \|A\|$, and

$$\|\text{diag}(M_j)_j\| = \max_j \|M_j\| \text{ for any set of matrices } M_j. \quad (2.7)$$

Fact 2.3. *If A_0 is a submatrix of a matrix A , then $\sigma_j(A) \geq \sigma_j(A_0)$ for all j .*

Proof. [GL96, Corollary 8.6.3] implies the claimed bound where A_0 is any block of columns of the matrix A . Transposition of a matrix and permutations of its rows and columns do not change singular values, and thus we can extend the bounds to all submatrices A_0 . \square

Theorem 2.1. *We have $|\sigma_j(C) - \sigma_j(C + E)| \leq \|E\|$ for all $m \times n$ matrices C and E and all j .*

Proof. See [GL96, Corollary 8.6.2] or [S98, Corollary 4.3.2]. \square

In Sections 7–9 we use the following definitions. For every integer k in the range $1 \leq k < \text{rank}(A)$ define the partition $S_A = (S_{k,A} \mid S_{A,m-k})$ and $T_A = (T_{k,A} \mid T_{A,n-k})$ where the submatrices $S_{k,A}$ and $T_{k,A}$ are formed by the first k columns of the matrices S_A and T_A , respectively. Write $\Sigma_{k,A} = \text{diag}(\sigma_j(A))_{j=1}^k$, $\mathbb{S}_{k,A} = \mathcal{R}(S_{k,A})$ and $\mathbb{T}_{k,A} = \mathcal{R}(T_{k,A})$. If $\sigma_k > \sigma_{k+1}$, then $\mathbb{S}_{k,A}$ and $\mathbb{T}_{k,A}$ are the left and right *leading singular spaces*, respectively, associated with the k largest singular values of the matrix A , whereas their orthogonal complements $\mathbb{S}_{A,m-k} = \mathcal{R}(S_{A,m-k})$ and $\mathbb{T}_{A,n-k} = \mathcal{R}(T_{A,n-k})$ are the left and right *trailing singular spaces*, respectively, associated with the other singular values of A . The pairs of subscripts $\{k, A\}$ versus $\{A, m-k\}$ and $\{A, n-k\}$ mark the leading versus trailing singular spaces. The left singular spaces of A are the right singular spaces of A^T and vice versa. All matrix bases for the singular spaces $\mathbb{S}_{k,A}$ and $\mathbb{T}_{k,A}$ are given by matrices $S_{k,A}X$ and $T_{k,A}Y$, respectively, for nonsingular $k \times k$ matrices X and Y . Orthogonal matrices X and Y define orthogonal matrix bases for these spaces. B is an *approximate matrix basis* for a space \mathbb{S} within a relative error norm bound τ if there exists a matrix E such that $B + E$ is a matrix basis for this space \mathbb{S} and if $\|E\| \leq \tau\|B\|$.

2.4 Inverses, generalized inverses, and perturbation bounds

$A^+ = T_A \text{diag}(\widehat{\Sigma}_A^{-1}, O_{n-\rho, m-\rho}) S_A^T$ is the Moore–Penrose pseudo-inverse of the matrix A of (2.5), and

$$\|A^+\| = 1/\sigma_\rho(A) \quad (2.8)$$

for a matrix A of a rank ρ . A^{+T} stands for $(A^+)^T = (A^T)^+$, and A^{-T} stands for $(A^{-1})^T = (A^T)^{-1}$.

An $n \times m$ matrix $X = A^{(I)}$ is a left inverse of an $m \times n$ matrix A if $XA = I$ and is its right inverse if $AX = I$. A^+ is a left or right inverse $A^{(I)}$ if and only if a matrix A has full rank. $A^{(I)}$ is unique and is equal to A^{-1} if A is a nonsingular matrix. Theorem 2.1 implies the following bound.

Theorem 2.2. *Suppose two matrices $C, C + E \in \mathbb{C}^{m \times n}$ have full rank. Then $\|(C + E)^+ - C^+\| \leq \|E\| \|(C + E)^+ C^+\|$.*

This bound can be improved where the matrices C and $C + E$ are nonsingular.

Theorem 2.3. *Suppose C and $C + E$ are two nonsingular matrices of the same size and $\|C^{-1}E\| = \theta < 1$. Then $\|I - (C + E)^{-1}C\| \leq \frac{\theta}{1-\theta}$ and $\|(C + E)^{-1} - C^{-1}\| \leq \frac{\theta}{1-\theta}\|C^{-1}\|$, in particular $\|(C + E)^{-1} - C^{-1}\| \leq 0.5\|C^{-1}\|$ if $\theta \leq 1/3$.*

Proof. See [S98, Corollary 1.4.19] for $P = -C^{-1}E$. □

2.5 SMW and dual SMW formulae

Theorem 2.4. [GL96, page 50], [S98, Corollary 4.3.2]. *Suppose that $U, V \in \mathbb{R}^{n \times r}$, the matrices $A \in \mathbb{R}^{n \times n}$ and $C = A + UV^T$ are nonsingular, and $0 < r < n$. Then the matrix $G = I_r - V^T C^{-1} U$ is nonsingular and we have the Sherman–Morrison–Woodbury (hereafter SMW) formula*

$$A^{-1} = C^{-1} + C^{-1} U G^{-1} V^T C^{-1}.$$

Corollary 2.1. *Suppose that $U_-, V_- \in \mathbb{R}^{n \times q}$, $A \in \mathbb{R}^{n \times n}$, A and $A^{-1} + U_- V_-^T$ are nonsingular matrices, and $0 < q < n$. Write*

$$C_-^{-1} = A^{-1} + U_- V_-^T, \quad H = I_q + V_-^T A U_-, \quad (2.9)$$

Then the matrix H is nonsingular and the following dual SMW formula holds,

$$C_- = A - A U_- H^{-1} V_-^T A. \quad (2.10)$$

Proof. Apply Theorem 2.4 to the matrices A^{-1} , U_- , V_- and C_-^{-1} replacing the matrices A , U , V and C , respectively. □

2.6 Condition number, numerical rank and numerical nullity, generic conditioning profile

$\kappa(A) = \frac{\sigma_1(A)}{\sigma_\rho(A)} = \|A\| \|A^+\|$ is the condition number of an $m \times n$ matrix A of a rank ρ . Such matrix is *ill conditioned* if $\sigma_1(A) \gg \sigma_\rho(A)$ and is *well conditioned* otherwise. See [D83], [GL96, Sections 2.3.2, 2.3.3, 3.5.4, 12.5], [H02, Chapter 15], [KL94], and [S98, Section 5.3] on the estimation of norms and condition numbers of nonsingular matrices.

An $m \times n$ matrix A has *numerical rank* $\text{nrnk}(A)$, not exceeding $\text{rank}(A)$, and has the right numerical nullity $\text{nnul}(A) = n - \text{nrnk}(A)$ or just *numerical nullity* if the ratios $\sigma_j(A)/\|A\|$ are small for $j > \text{nrnk}(A)$ but not for $j \leq \text{nrnk}(A)$. The *left numerical nullity* of the matrix A equals the numerical nullity $\text{nnul}(A^T) = m - \text{nrnk}(A)$ of the $n \times m$ transpose A^T and coincides with the numerical nullity of A if and only if $m = n$.

Remark 2.1. One can specify the adjective “small” above as “smaller than a fixed positive tolerance”. The choice of the tolerance can be a challenge, e.g., for the matrix $\text{diag}(1.1^{-j})_{j=0}^{999}$.

If a well conditioned $m \times n$ matrix A has a rank $\rho < l = \min\{m, n\}$, then almost all its close neighbours have full rank l (see Section 3.2), and all of them have numerical rank ρ . Conversely, suppose a matrix A has a positive numerical rank $\rho = \text{nrnk}(A)$ and *truncate its SVD* by setting to 0 all its singular values, except for the ρ largest ones. Then the resulting matrix $A - E$ is well conditioned and has rank ρ and $\|E\| = \sigma_{\rho+1}(A)$, and so $A - E$ is a rank- ρ approximation to the matrix A within the error norm bound $\sigma_{\rho+1}(A)$. At a lower computational cost we can obtain rank- ρ approximations of the matrix A from its rank-revealing factorizations [GE96], [HP92], [P00a], and we further decrease the computational cost by applying randomized algorithms in Sections 7 and 8.

An $m \times n$ matrix has *generic conditioning profile* (cf. the end of Section 2.2) if it has a numerical rank ρ and if its leading $i \times i$ blocks are nonsingular and well conditioned for $i = 1, \dots, \rho$. If such matrix has full rank (that is if $\rho = \min\{m, n\}$) and if it is well conditioned itself, then we call it *strongly well conditioned*. The following theorem shows that GENP and block Gaussian elimination applied to a strongly well conditioned matrix are numerically safe.

Theorem 2.5. Cf. [PQZa, Theorem 5.1]. Assume GENP or block Gaussian elimination applied to an $n \times n$ matrix A and write $N = \|A\|$ and $N_- = \max_{j=1}^n \|(A_j^{(j)})^{-1}\|$. Then the absolute values of all pivot elements of GENP and the norms of all pivot blocks of block Gaussian elimination do not exceed $N + N_- N^2$, whereas the absolute values of the reciprocals of these elements and the norms of the inverses of the blocks do not exceed N_- .

2.7 Toeplitz, Hankel and f -circulant matrices

A *Toeplitz* $m \times n$ matrix $T_{m,n} = (t_{i-j})_{i,j=1}^{m,n}$ is defined by its first row and column, that is by the vector $(t_h)_{h=1-n}^{m-1}$ of dimension $m+n-1$. We write $T_n = T_{n,n} = (t_{i-j})_{i,j=1}^{n,n}$ (see (2.11)).

A lower *triangular Toeplitz* $n \times n$ matrix $Z(\mathbf{t}) = (t_{i-j})_{i,j=1}^n$ (where $t_k = 0$ for $k < 0$) is defined by its first column $\mathbf{t} = (t_h)_{h=0}^{n-1}$. We write $Z(\mathbf{t})^T = (Z(\mathbf{t}))^T$. $Z = Z_0 = Z(\mathbf{e}_2)$ is the downshift $n \times n$ (see (2.11)). We have $Z\mathbf{v} = (v_i)_{i=0}^{n-1}$ and $Z(\mathbf{v}) = Z_0(\mathbf{v}) = \sum_{i=1}^n v_i Z^{i-1}$ for $\mathbf{v} = (v_i)_{i=1}^n$ and $v_0 = 0$,

$$T_n = \begin{pmatrix} t_0 & t_{-1} & \cdots & t_{1-n} \\ t_1 & t_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{-1} \\ t_{n-1} & \cdots & t_1 & t_0 \end{pmatrix}, \quad Z = \begin{pmatrix} 0 & \cdots & 0 \\ 1 & \ddots & & \\ \vdots & \ddots & \ddots & \vdots \\ & \ddots & 0 & \\ 0 & \cdots & 1 & 0 \end{pmatrix}. \quad (2.11)$$

Combine the equations $\|Z(\mathbf{v})\|_1 = \|Z(\mathbf{v})\|_\infty = \|\mathbf{v}\|_1$ with (2.2) to obtain

$$\|Z(\mathbf{v})\| \leq \|\mathbf{v}\|_1. \quad (2.12)$$

Theorem 2.6. Write $T_k = (t_{i-j})_{i,j=0}^{k-1}$ for $k = n, n+1$.

(a) Let the matrix T_n be nonsingular and write $\mathbf{p} = T_n^{-1}\mathbf{e}_1$ and $\mathbf{q} = T_n^{-1}\mathbf{e}_n$. If $p_1 = \mathbf{e}_1^T \mathbf{p} \neq 0$, then $p_1 T_n^{-1} = Z(\mathbf{p})Z(J\mathbf{q})^T - Z(Z\mathbf{q})Z(ZJ\mathbf{p})^T$.

In parts (b) and (c) below let the matrix T_{n+1} be nonsingular and write $\hat{\mathbf{v}} = (v_i)_{i=0}^n = T_{n+1}^{-1}\mathbf{e}_1$, $\mathbf{v} = (v_i)_{i=0}^{n-1}$, $\mathbf{v}' = (v_i)_{i=1}^n$, $\hat{\mathbf{w}} = (w_i)_{i=0}^n = T_{n+1}^{-1}\mathbf{e}_{n+1}$, $\mathbf{w} = (w_i)_{i=0}^{n-1}$, and $\mathbf{w}' = (w_i)_{i=1}^n$.

(b) If $v_0 \neq 0$, then the matrix T_n is nonsingular and $v_0 T_n^{-1} = Z(\mathbf{v})Z(J\mathbf{w}')^T - Z(\mathbf{w})Z(J\mathbf{v}')^T$.

(c) If $v_n \neq 0$, then the matrix $T_{1,0} = (t_{i-j})_{i=1,j=0}^{n,n-1}$ is nonsingular and $v_n T_{1,0}^{-1} = Z(\mathbf{w})Z(J\mathbf{v}')^T - Z(\mathbf{v})Z(J\mathbf{w}')^T$.

Proof. See [GS72] on parts (a) and (b); see [GK72] on part (c). \square

$Z_f = Z + f\mathbf{e}_1^T \mathbf{e}_n$ for a scalar $f \neq 0$ denotes the $n \times n$ matrix of f -circular shift. An f -circulant matrix $Z_f(\mathbf{v}) = \sum_{i=1}^n v_i Z_f^{i-1}$ is a special Toeplitz $n \times n$ matrix defined by its first column vector $\mathbf{v} = (v_i)_{i=1}^n$ and a scalar f . f -circulant matrix is called *circulant* if $f = 1$ and *skew circulant* if $f = -1$. By replacing f with 0 we arrive at a lower triangular Toeplitz matrix $Z(\mathbf{v})$. The following theorem implies that the inverses (wherever they are defined) and pairwise products of f -circulant $n \times n$ matrices are f -circulant and can be computed in $O(n \log n)$ flops.

Theorem 2.7. (See [CPW74].) We have $Z_1(\mathbf{v}) = \Omega^{-1}D(\Omega\mathbf{v})\Omega$. More generally, for any $f \neq 0$, we have $Z_{f^n}(\mathbf{v}) = U_f^{-1}D(U_f\mathbf{v})U_f$ where $U_f = \Omega D(\mathbf{f})$, $\mathbf{f} = (f^i)_{i=0}^{n-1}$, $D(\mathbf{u}) = \text{diag}(u_i)_{i=0}^{n-1}$ for a vector $\mathbf{u} = (u_i)_{i=0}^{n-1}$, $\Omega = (\omega_n^{ij})_{i,j=0}^{n-1}$ is the $n \times n$ matrix of the discrete Fourier transform at n points, $\omega_n = \exp(\frac{2\pi}{n}\sqrt{-1})$ being a primitive n -th root of 1, and $\Omega^{-1} = \frac{1}{n}(\omega_n^{-ij})_{i,j=0}^{n-1} = \frac{1}{n}\Omega^H$.

Hankel $m \times n$ matrices $H = (h_{i+j})_{i,j=1}^{m,n}$ can be defined equivalently as the products $H = TJ_n$ or $H = J_m T$ of $m \times n$ Toeplitz matrices T and the Hankel reflection matrices $J = J_m$ or J_n . Note that $J = J^{-1} = J^T$ and obtain the following simple fact.

Fact 2.4. For $m = n$ we have $T = HJ$, $H^{-1} = JT^{-1}$ and $T^{-1} = JH^{-1}$ if $H = TJ$, whereas $T = JH$, $H^{-1} = JT^{-1}$ and $T^{-1} = H^{-1}J$ if $H = JT$. Furthermore in both cases $\kappa(H) = \kappa(T)$.

By using the equations above we can readily extend any Toeplitz matrix inversion algorithm to Hankel matrix inversion and vice versa, preserving the flop count and condition numbers. E.g. $(JT)^{-1} = T^{-1}J$, $(TJ)^{-1} = JT^{-1}$, $(JH)^{-1} = H^{-1}J$ and $(HJ)^{-1} = JH^{-1}$.

2.8 Toeplitz-like, Hankel-like and some other structured matrices

Let us extend the class of Toeplitz and Hankel matrices to a more general class of structured matrices, which we only employ in Section 8. With every pair of $n \times n$ operator matrices A and B associate the class of $n \times n$ matrices M for which the rank of the Sylvester *displacement* $AM - MB$ (called the *displacement rank* of M) is small in context. The matrices T with the structure of Toeplitz type (we call them *Toeplitz-like* matrices) have small *displacement ranks* $d = d(A, B)$ for $A = Z_e$ and $B = Z_f$ and for any pair of scalars e and f . Such matrices extend the class of Toeplitz matrices, for which $d \leq 2$. Any variation of a pair of scalars e and f can change the displacement rank of a matrix by at most 2, and so the class of Toeplitz-like matrices is independent of the choice of such pair.

Every matrix of a rank d , and in particular a displacement of a rank d , can be nonuniquely represented as the sum of d outer products $\mathbf{g}_j \mathbf{h}_j^T$ of d pairs of vectors \mathbf{g}_j and \mathbf{h}_j for $j = 1, \dots, d$. Motivated by the following result we call the pair of matrices $G = G_{Z_e, Z_f}(M) = (\mathbf{g}_j)_{j=1}^d$ and $H = H_{Z_e, Z_f}(M) = (\mathbf{h}_j)_{j=1}^d$, made up of the vectors \mathbf{g}_j and \mathbf{h}_j , a *displacement generator* of length d for the matrix M and for the operator matrices Z_e and Z_f where $e \neq f$ (cf. [P01, Example 4.4.2]).

Theorem 2.8. If $Z_e M - M Z_f = \sum_{j=1}^d \mathbf{g}_j \mathbf{h}_j^T$ for a pair of distinct scalars e and f , then

$$(e - f)M = \sum_{j=1}^d Z_e(\mathbf{g}_j)Z_f(J\mathbf{h}_j). \quad (2.13)$$

The theorem expresses the matrix through the displacement generator $\{G, H\}$ by using $2dn$ parameters instead of n^2 entries. For $d \ll n$, this is a dramatic compression, which furthermore reduces multiplication of the matrix M by a vector essentially to $2d$ multiplications of circulant matrices by vectors, that is to $O(dn \log n)$ flops. Moreover we can operate with matrices by using their displacement representation, which preserves Toeplitz-like structure and can accelerate the computations dramatically where $d \ll n$. For Toeplitz-like matrices T , T_1 and T_2 , scalars e , f , α , and β , and operator matrices $A = Z_e$, $B = Z_f$, and $C = Z_c$, we can readily obtain the Toeplitz-like matrices T^{-1} (if the matrix T is nonsingular), T^T , $\alpha T_1 + \beta T_2$, and $T_1 T_2$. The following theorem bounds the growth of the length of the associated displacement generators and the respective flop cost.

Theorem 2.9. *Assume that $n \times n$ matrices T , T_1 , and T_2 have been represented with their displacement generators of lengths d_1 , d_2 , and d , respectively, for appropriate operator matrices $A = Z_e$ and $B = Z_f$, defining Toeplitz-like structure. Then there exist displacement generators of length d for T^{-1} (provided that the matrix T is nonsingular) and T^T , of length $d_1 + d_2$ for $\alpha T_1 + \beta T_2$, and of length $d_1 + d_2 + O(1)$ for $T_1 T_2$ (for appropriate operator matrices defining Toeplitz-like structure and for any pair of scalar α and β). One can compute these generators by using $O(d^2 n \log^2 n)$ and $O(d_1 d_2 n \log n)$ flops, respectively.*

Proof. The theorem readily follows from Theorem A.1 and Corollary A.1 in Appendix A, which also define all the respective displacement generators. \square

A matrix H is *Hankel-like* if $\text{rank}(AH - HB)$ is small where $A = Z_e$ and $B = Z_f^T$ for two scalars e and f or alternatively where $A = Z_e^T$ and $B = Z_f$. It follows that MN is a Hankel-like matrix if one of the factors is a Toeplitz-like matrix and another is a Hankel-like matrix, whereas MN is a Toeplitz-like matrix if both M and N are Hankel-like matrices or both are Toeplitz-like matrices. We can alternatively define Hankel-like matrices as the products TJ or JT where T is a Toeplitz-like matrix, or we can define Toeplitz-like matrices T as the products HJ and JH where H are Hankel-like matrices (cf. Fact 2.4). By using these properties we can readily extend our algorithms as well as expressions (2.13) (cf. [P01, Example 4.4.4]) from the case of Toeplitz and Toeplitz-like to Hankel and Hankel-like matrices, preserving the flop count.

Remark 2.2. *By choosing the operator matrices A and B among f -circulant and appropriate diagonal matrices we define the important classes of matrices M with the structures of Vandermonde and Cauchy types whose displacement rank, that is $\text{rank}(AM - MB)$, is small. This extends the classes of Vandermonde matrices $V_{\mathbf{x}} = (x_i^{j-1})_{i,j=1}^n$, having displacement rank 1 for the operator matrices $A = \text{diag}(x_i)_{i=1}^n$ and $B = Z_f^T$ for a scalar f , and Cauchy matrices $C_{\mathbf{s}, \mathbf{t}} = (\frac{1}{s_i - t_j})_{i,j=1}^n$, having displacement rank 1 for the operator matrices $A = \text{diag}(s_i)_{i=1}^n$ and $B = \text{diag}(t_j)_{j=1}^n$. Alternatively [P90], [P01], the matrices of these classes can be defined as the products UMV where M is a Toeplitz matrix, whereas U and V are properly selected among Vandermonde matrices, their transposes and the identity matrices. Similarly to the Toeplitz–Hankel link at the end of the previous subsection, this enables us to extend any successful algorithm for Cauchy-like inversion to Toeplitz-like, Hankel-like and Vandermonde-like inversion and vice versa because $(UMV)^{-1} = V^{-1}M^{-1}U^{-1}$ [P90], although unlike the orthogonal reversion matrix J , Vandermonde multipliers and their transposes are usually ill conditioned except for a narrow but important class including the matrices Ω and Ω^H of Theorem 2.7. Theorems 2.8 and 2.9 and other basic properties of Toeplitz-like and Hankel-like matrices can be extended to the matrices having structures of Vandermonde or Cauchy types (see [P00], [P01] or Appendix A).*

3 Ranks and conditioning of Gaussian random matrices

3.1 Random variables and Gaussian random matrices

Definition 3.1. $F_\gamma(y) = \text{Probability}\{\gamma \leq y\}$ for a real random variable γ is the cumulative distribution function (cdf) of γ evaluated at y . $F_{g(\mu, \sigma)}(y) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^y \exp(-\frac{(x-\mu)^2}{2\sigma^2}) dx$ for a Gaussian

random variable $g(\mu, \sigma)$ with a mean μ and a positive variance σ^2 , and so

$$\mu - 4\sigma \leq y \leq \mu + 4\sigma \text{ with a probability near } 1. \quad (3.1)$$

Definition 3.2. A matrix or a vector is a Gaussian random matrix or vector with a mean μ and a positive variance σ^2 if it is filled with independent identically distributed Gaussian random variables, all having the mean μ and variance σ^2 . $\mathcal{G}_{\mu, \sigma}^{m \times n}$ is the set of such Gaussian random $m \times n$ matrices (which are standard for $\mu = 0$ and $\sigma^2 = 1$). By restricting this set to Toeplitz or f -circulant matrices we obtain the sets $\mathcal{T}_{\mu, \sigma}^{m \times n}$ and $\mathcal{Z}_{f, \mu, \sigma}^{n \times n}$ of Gaussian random Toeplitz and Gaussian random f -circulant matrices, respectively.

Definition 3.3. $\chi_{\mu, \sigma, n}(y)$ is the cdf of the norm $\|\mathbf{v}\| = (\sum_{i=1}^n v_i^2)^{1/2}$ of a Gaussian random vector $\mathbf{v} = (v_i)_{i=1}^n \in \mathcal{G}_{\mu, \sigma}^{n \times 1}$. For $y \geq 0$ we have $\chi_{0,1,n}(y) = \frac{2}{2^{n/2}\Gamma(n/2)} \int_0^y x^{n-1} \exp(-x^2/2) dx$ where $\Gamma(h) = \int_0^\infty x^{h-1} \exp(-x) dx$, $\Gamma(n+1) = n!$ for nonnegative integers n .

3.2 Nondegeneration of Gaussian random matrices

The total degree of a multivariate monomial is the sum of its degrees in all its variables. The total degree of a polynomial is the maximal total degree of its monomials.

Lemma 3.1. [DL78], [S80], [Z79]. For a set Δ of a cardinality $|\Delta|$ in any fixed ring let a polynomial in m variables have a total degree d and let it not vanish identically on this set. Then the polynomial vanishes in at most $d|\Delta|^{m-1}$ points.

We assume that Gaussian random variables range over infinite sets Δ , usually over the real line or its interval. Then the lemma implies that a nonzero polynomial vanishes with probability 0. Consequently a square Gaussian random general, Toeplitz or circulant matrix is nonsingular with probability 1 because its determinant is a polynomial in the entries. Likewise rectangular Gaussian random general, Toeplitz and circulant matrices have full rank with probability 1. Furthermore all entries of such matrix A and of its adjoint $\text{adj } A$ are subdeterminants and thus are nonzero with probability 1. Clearly this property of the adjoint also holds for the inverse $A^{-1} = \frac{\text{adj } A}{\det A}$ if the matrix A is nonsingular. Hereafter, wherever this causes no confusion, we assume by default that Gaussian random general, Toeplitz and circulant matrices have full rank, and their inverses (if defined) have nonzero entries. These properties can be readily extended to the products of the latter matrices by nonsingular and orthogonal matrices, and further to various functions of general, sparse and structured matrices. Moreover similar properties hold with probability near 1 where the random variables are sampled under the uniform probability distribution from a finite set of a large cardinality (see Appendix A).

3.3 Extremal singular values of Gaussian random matrices

Besides having full rank with probability 1, Gaussian random matrices in Definition 3.2 are expected to be well conditioned [D88], [E88], [ES05], [CD05], [B11], and even the sum $M + A$ for $M \in \mathbb{R}^{m \times n}$ and $A \in \mathcal{G}_{\mu, \sigma}^{m \times n}$ is expected to be well conditioned unless the ratio $\sigma/\|M\|$ is small or large [SST06].

The following theorem states an upper bound proportional to y on the cdf $F_{1/\|A^+\|}(y)$, that is on the probability that the smallest positive singular value $1/\|A^+\| = \sigma_l(A)$ of a Gaussian random matrix A is less than a nonnegative scalar y (cf. (2.8)) and consequently on the probability that the norm $\|A^+\|$ exceeds a positive scalar x . The stated bound still holds if we replace the matrix A by $A - B$ for any fixed matrix B , although for $B = O_{m,n}$ the bounds can actually be strengthened by a factor $y^{|m-n|}$ [ES05], [CD05].

Theorem 3.1. Suppose $A \in \mathcal{G}_{\mu, \sigma}^{m \times n}$, $B \in \mathbb{R}^{m \times n}$, $l = \min\{m, n\}$, $x > 0$, and $y \geq 0$. Then $F_{\sigma_l(A-B)}(y) \leq 2.35 \sqrt{l}y/\sigma$, that is $\text{Probability}\{\|(A - B)^+\| \geq 2.35x\sqrt{l}/\sigma\} \leq 1/x$.

Proof. For $m = n$ this is [SST06, Theorem 3.3]. Apply Fact 2.3 to extend it to any pair $\{m, n\}$. \square

The following two theorems supply lower bounds $F_{\|A\|}(z)$ and $F_{\kappa(A)}(y)$ on the probabilities that $\|A\| \leq z$ and $\kappa(A) \leq y$ for two scalars y and z , respectively, and a Gaussian random matrix A . We do not use the second theorem, but state it for the sake of completeness and only for square $n \times n$ matrices A . The theorems imply that the functions $1 - F_{\|A\|}(z)$ and $1 - F_{\kappa(A)}(y)$ decay as $z \rightarrow \infty$ and $y \rightarrow \infty$, respectively, and that the decays are exponential in $-z^2$ and proportional to $\sqrt{\log y}/y$, respectively. For small values $y\sigma$ and a fixed n the lower bound of Theorem 3.3 becomes negative, in which case the theorem becomes trivial. Unlike Theorem 3.1, in both theorems we assume that $\mu = 0$.

Theorem 3.2. [DS01, Theorem II.7]. Suppose $A \in \mathcal{G}_{0,\sigma}^{m \times n}$, $h = \max\{m, n\}$ and $z \geq 2\sigma\sqrt{h}$. Then $F_{\|A\|}(z) \geq 1 - \exp(-(z - 2\sigma\sqrt{h})^2/(2\sigma^2))$, and so the norm $\|A\|$ is expected to have order $\sigma\sqrt{h}$.

Theorem 3.3. [SST06, Theorem 3.1]. Suppose $0 < \sigma \leq 1$, $y \geq 1$, $A \in \mathcal{G}_{0,\sigma}^{n \times n}$. Then the matrix A has full rank with probability 1 and $F_{\kappa(A)}(y) \geq 1 - (14.1 + 4.7\sqrt{(2 \ln y)/n})n/(y\sigma)$.

Proof. See [SST06, the proof of Lemma 3.2]. \square

3.4 Extremal singular values of Gaussian random Toeplitz matrices

A matrix $T_n = (t_{i-j})_{i,j=1}^n$ is the sum of two triangular Toeplitz matrices

$$T_n = Z(\mathbf{t}) + Z(\mathbf{t}_-)^T, \quad \mathbf{t} = (t_i)_{i=0}^{n-1}, \quad \mathbf{t}_- = (t'_{-i})_{i=0}^{n-1}, \quad t'_0 = 0. \quad (3.2)$$

If $T_n \in \mathcal{T}_{\mu,\sigma}^{n \times n}$, then T_n has $2n - 1$ pairwise independent entries in $\mathcal{G}_{\mu,\sigma}$. Thus (2.12) implies that

$$\|T_n\| \leq \|Z(\mathbf{t})\| + \|Z(\mathbf{t}_-)^T\| \leq \|\mathbf{t}\|_1 + \|\mathbf{t}_-\|_1 = \|(t_i)_{i=1-n}^{n-1}\|_1 \leq \sqrt{2n} \|(t_i)_{i=1-n}^{n-1}\|.$$

Recall Definition 3.2 and obtain

$$F_{\|T_n\|}(y) \geq \chi_{\mu,\sigma,2n-1}(y/\sqrt{2n}). \quad (3.3)$$

Next we estimate the norm $\|T_n^{-1}\|$ for $T_n \in \mathcal{T}_{\mu,\sigma}^{n \times n}$.

Lemma 3.2. [SST06, Lemma A.2]. For a nonnegative scalar y , a unit vector $\mathbf{t} \in \mathbb{R}^{n \times 1}$, and a vector $\mathbf{b} \in \mathcal{G}_{\mu,\sigma}^{n \times 1}$, we have $F_{|\mathbf{t}^T \mathbf{b}|}(y) \leq \sqrt{\frac{2}{\pi}} \frac{y}{\sigma}$.

Remark 3.1. The latter bound is independent of μ and n ; it holds for any μ even if all coordinates of the vector \mathbf{b} are fixed except for a single coordinate in $\mathcal{G}_{\mu,\sigma}$.

Theorem 3.4. Given a matrix $T_n = (t_{i-j})_{i,j=1}^n \in \mathcal{T}_{\mu,\sigma}^{n \times n}$, assumed to be nonsingular (cf. Section 3.2), write $p_1 = \mathbf{e}_1^T T_n^{-1} \mathbf{e}_1$. Then $F_{1/\|p_1 T_n^{-1}\|}(y) \leq 2n\alpha\beta$ for two random variables α and β such that

$$F_\alpha(y) \leq \sqrt{\frac{2n}{\pi}} \frac{y}{\sigma} \quad \text{and} \quad F_\beta(y) \leq \sqrt{\frac{2n}{\pi}} \frac{y}{\sigma} \quad \text{for } y \geq 0. \quad (3.4)$$

Proof. Recall from part (a) of Theorem 2.6 that $p_1 T_n^{-1} = Z(\mathbf{p})Z(J\mathbf{q})^T - Z(Z\mathbf{q})Z(ZJ\mathbf{p})^T$. Therefore $\|p_1 T_n^{-1}\| \leq \|Z(\mathbf{p})\| \|Z(J\mathbf{q})^T\| + \|Z(Z\mathbf{q})\| \|Z(ZJ\mathbf{p})^T\|$ for $\mathbf{p} = T_n^{-1} \mathbf{e}_1$, $\mathbf{q} = T_n^{-1} \mathbf{e}_n$, and $p_1 = \mathbf{p}^T \mathbf{e}_1$. It follows that $\|p_1 T_n^{-1}\| \leq \|Z(\mathbf{p})\| \|Z(J\mathbf{q})\| + \|Z(Z\mathbf{q})\| \|Z(ZJ\mathbf{p})\|$ since $\|A\| = \|A^T\|$ for all matrices A . Furthermore $\|p_1 T_n^{-1}\| \leq \|\mathbf{p}\|_1 \|J\mathbf{q}\|_1 + \|Z\mathbf{q}\|_1 \|ZJ\mathbf{p}\|_1$ due to (2.12). Clearly $\|J\mathbf{v}\|_1 = \|\mathbf{v}\|_1$ and $\|Z\mathbf{v}\|_1 \leq \|\mathbf{v}\|_1$ for every vector \mathbf{v} , and so (cf. (2.2))

$$\|p_1 T_n^{-1}\| \leq 2\|\mathbf{p}\|_1 \|\mathbf{q}\|_1 \leq 2n\|\mathbf{p}\| \|\mathbf{q}\|. \quad (3.5)$$

By definition the vector \mathbf{p} is orthogonal to the vectors $T_n \mathbf{e}_2, \dots, T_n \mathbf{e}_n$, whereas $\mathbf{p}^T T_n \mathbf{e}_1 = 1$ (cf. [SST06]). Consequently the vectors $T_n \mathbf{e}_2, \dots, T_n \mathbf{e}_n$ uniquely define the vector $\mathbf{u} = \mathbf{p}/\|\mathbf{p}\|$, whereas $|\mathbf{u}^T T_n \mathbf{e}_1| = 1/\|\mathbf{p}\|$. The last coordinate t_{n-1} of the vector $T_n \mathbf{e}_1$ is independent of the vectors $T_n \mathbf{e}_2, \dots, T_n \mathbf{e}_n$ and consequently of the vector \mathbf{u} . Apply Remark 3.1 to estimate the cdf of the random variable $\alpha = 1/\|\mathbf{p}\| = |\mathbf{u}^T T_n \mathbf{e}_1|$ and obtain that $F_\alpha(y) \leq \sqrt{\frac{2n}{\pi}} \frac{y}{\sigma}$ for $y \geq 0$.

Likewise the $n - 1$ column vectors $T\mathbf{e}_1, \dots, T_{n-1}$ define the vector $\mathbf{v} = \beta \mathbf{q}$ for $\beta = 1/\|\mathbf{q}\| = |\mathbf{v}^T T_n \mathbf{e}_n|$. The first coordinate t_{1-n} of the vector $T_n \mathbf{e}_n$ is independent of the vectors $T\mathbf{e}_1, \dots, T_{n-1}$ and consequently of the vector \mathbf{v} . Apply Remark 3.1 to estimate the cdf of the random variable β and obtain that $F_\beta(y) \leq \sqrt{\frac{2n}{\pi}} \frac{y}{\sigma}$ for $y \geq 0$. Finally combine these bounds on the cdfs $F_\alpha(y)$ and $F_\beta(y)$ with (3.5). \square

By applying parts (b) and (c) of Theorem 2.6 instead of its part (a), we similarly deduce the bounds $\|v_0 T_{n+1}^{-1}\| \leq 2\alpha\beta$ and $\|v_n T_{n+1}^{-1}\| \leq 2\alpha\beta$ for two pairs of random variables α and β that satisfy (3.4) for $n + 1$ replacing n . We have $p_1 = \frac{\det T_{n-1}}{\det T_n}$, $v_0 = \frac{\det T_n}{\det T_{n+1}}$, and $v_n = \frac{\det T_{0,1}}{\det T_{n+1}}$ for $T_{0,1} = (t_{i-j})_{i=0, j=1}^{n-1, n}$. Next we bound the geometric means of the ratios $|\frac{\det T_{h+1}}{\det T_h}|$ for $h = 1, \dots, k - 1$. $1/|p_1|$ and $1/|v_0|$ are such ratios for $k = n - 1$ and $k = n$, respectively, whereas the ratio $1/|v_n|$ is similar to $1/|v_0|$, under slightly distinct notation.

Theorem 3.5. *Let $T_h \neq O$ denote $h \times h$ matrices for $h = 1, \dots, k$ whose entries have absolute values at most t for a fixed scalar or random variable t , e.g. for $t = \|T\|$. Furthermore let $T_1 = (t)$. Then the geometric mean $(\prod_{h=1}^{k-1} |\frac{\det T_{h+1}}{\det T_h}|)^{1/(k-1)} = \frac{1}{t} |\det T_k|^{1/(k-1)}$ is at most $k^{\frac{1}{2}(1+\frac{1}{k-1})} t$.*

Proof. The theorem follows from Hadamard's upper bound $|\det M| \leq k^{k/2} t^k$, which holds for any $k \times k$ matrix $M = (m_{i,j})_{i,j=1}^k$ with $\max_{i,j=1}^k |m_{i,j}| \leq t$. \square

The theorem says that the geometric mean of the ratios $|\det T_{h+1}/\det T_h|$ for $h = 1, \dots, k - 1$ is not greater than $k^{0.5+\epsilon(k)} t$ where $\epsilon(k) \rightarrow 0$ as $k \rightarrow \infty$. Furthermore if $T_n \in \mathcal{G}_{\mu, \sigma}^{n \times n}$ we can write $t = \|T\|$ and apply (3.3) to bound the cdf of t .

3.5 Extremal singular values of Gaussian random circulant matrices

Next we estimate the norms of a random Gaussian f -circulant matrix and its inverse.

Theorem 3.6. *Assume $y \geq 0$ and a circulant $n \times n$ matrix $T = Z_1(\mathbf{v})$ for $\mathbf{v} \in \mathcal{G}_{\mu, \sigma}^{n \times 1}$. Then*

$$(a) F_{\|T\|}(y) \geq \chi_{\mu, \sigma, n}(\sqrt{\frac{2}{n}} y) \text{ for } \chi_{\mu, \sigma, n}(y) \text{ in Definition 3.3 and } (b) F_{1/\|T^{-1}\|}(y) \leq \sqrt{\frac{2}{\pi}} \frac{ny}{\sigma}.$$

Proof. For the matrix $T = Z_1(\mathbf{v})$ we have both equation (3.2) and the bound $\|\mathbf{t}_-\|_1 \leq \|\mathbf{t}\|_1$, and so $\|T\|_1 \leq 2\|\mathbf{t}\|_1$. Now part (a) of the theorem follows similarly to (3.3). To prove part (b) recall Theorem 2.7 and write $B = \Omega T \Omega^{-1} = D(\mathbf{u})$, $\mathbf{u} = (u_i)_{i=0}^{n-1} = \Omega \mathbf{v}$. We have $\sigma_j(T) = \sigma_j(B)$ for all j because $\frac{1}{\sqrt{n}} \Omega$ and $\sqrt{n} \Omega^{-1}$ are unitary matrices. By combining the equations $u_i = \mathbf{e}_i^T \Omega \mathbf{v}$, the bounds $|\Re(\mathbf{e}_i^T \Omega)| \geq 1$ for all i , and Lemma 3.2, deduce that $F_{|\Re(u_i)|}(y) \leq \sqrt{\frac{2}{\pi}} \frac{y}{\sigma}$ for $i = 1, \dots, n$. We have $F_{\sigma_n(B)}(y) = F_{\min_i |u_i|}(y)$ because $B = \text{diag}(u_i)_{i=0}^{n-1}$, and clearly $|u_i| \geq |\Re(u_i)|$. \square

Remark 3.2. *Our extensive experiments suggest that the estimates of Theorem 3.6 are overly pessimistic (cf. Table 10.4).*

Combining Theorem 2.7 with minimax property (2.6) implies that

$$\frac{1}{g(f)} \sigma_j(Z_1(\mathbf{v})) \leq \sigma_j(Z_f(\mathbf{v})) \leq g(f) \sigma_j(Z_1(\mathbf{v}))$$

for all vectors \mathbf{v} , scalars $f \neq 0$, $g(f) = \max\{|f|^2, 1/|f|^2\}$, and $j = 1, \dots, n$. Thus we can readily extend the estimates of Theorem 3.6 to f -circulant matrices for $f \neq 0$. In particular Gaussian random f -circulant matrices tend to be well conditioned unless $f \approx 0$ or $1/f \approx 0$.

4 Condition numbers of randomized matrix products and generic preconditioning

Next we deduce probabilistic lower bounds on the smallest singular values of the products of fixed and random matrices. We begin with three lemmas. The first of them is obvious, the second easily follows from minimax property (2.6).

Lemma 4.1. $\sigma_j(SM) = \sigma_j(MT) = \sigma_j(M)$ for all j if S and T are square orthogonal matrices.

Lemma 4.2. Suppose $\Sigma = \text{diag}(\sigma_i)_{i=1}^n$, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$, $G \in \mathbb{R}^{r \times n}$, $H \in \mathbb{R}^{n \times r}$. Then $\sigma_j(G\Sigma) \geq \sigma_j(G)\sigma_n$, $\sigma_j(\Sigma H) \geq \sigma_j(H)\sigma_n$ for all j . If also $\sigma_n > 0$, then $\text{rank}(G\Sigma) = \text{rank}(G)$, $\text{rank}(\Sigma H) = \text{rank}(H)$.

We employ the following result in the proof of Corollary 6.1.

Corollary 4.1. We have $\kappa(AB) \leq \kappa(A)\kappa(B)$ if A or B is a nonsingular matrix.

Proof. Assume SVDs $A = S_A \Sigma_A T_A^T$ of (2.5). Then $\sigma_j(AB) = \sigma_j(S_A \Sigma_A T_A^T B) = \sigma_j(\Sigma_A \hat{B})$ where $\hat{B} = T_A^T B$. Let A and consequently Σ_A be nonsingular $n \times n$ matrices. Apply Lemma 4.2 and deduce that $\sigma_j(\Sigma_A \hat{B}) \geq \sigma_j(\hat{B})\sigma_n(A)$, whereas $\sigma_j(\hat{B}) = \sigma_j(B)$ for all j . We have $\rho = \text{rank}(AB) = \text{rank}(B) \leq n$. Combine the relationships above for $j = \rho$ and obtain that $\sigma_\rho(AB) = \sigma_\rho(\Sigma_A \hat{B}) \geq \sigma_\rho(\hat{B})\sigma_n(A) = \sigma_\rho(B)\sigma_n(A)$, and so $\sigma_\rho(AB) \geq \sigma_\rho(B)\sigma_n(A)$. Also note that $\|AB\| \leq \|A\| \|B\|$. Combine the latter bounds and obtain that $\kappa(AB) = \|AB\|/\sigma_\rho(AB) \leq \|A\| \|B\|/(\sigma_\rho(B)\sigma_n(A)) = \kappa(A)\kappa(B)$. Similarly prove the claimed bound where B is a nonsingular matrix. \square

Lemma 4.3. [SST06, Proposition 2.2]. Suppose $H \in \mathcal{G}_{\mu, \sigma}^{m \times n}$, $SS^T = S^T S = I_m$, $TT^T = T^T T = I_n$. Then $SH \in \mathcal{G}_{\mu, \sigma}^{m \times n}$ and $HT \in \mathcal{G}_{\mu, \sigma}^{m \times n}$.

The following theorem implies that multiplication by standard Gaussian random matrix is unlikely to decrease the smallest positive singular value of a matrix dramatically, even though $UV = O$ for some pairs of rectangular orthogonal matrices U and V .

Theorem 4.1. Suppose $G' \in \mathcal{G}_{\mu, \sigma}^{r \times m}$, $H' \in \mathcal{G}_{\mu, \sigma}^{n \times r}$, $M \in \mathbb{R}^{m \times n}$, $G = G' + U$, $H = H' + V$ for some matrices U and V , $r(M) = \text{rank}(M)$, $x > 0$ and $y \geq 0$. Then $F_{1/\|(GM)+\|}(y) \leq F(y, M, \sigma)$ and $F_{1/\|(MH)+\|}(y) \leq F(y, M, \sigma)$ for $F(y, M, \sigma) = 2.35y\sqrt{\hat{r}}\|M^+\|/\sigma$ and $\hat{r} = \min\{r, r(M)\}$, that is $\text{Probability}\{|P^+| \geq 2.35x\sqrt{\hat{r}}\|M^+\|/\sigma\} \leq 1/x$ for $P = GM$ and $P = MH$.

Proof. With probability 1, the matrix MH has rank \hat{r} because $H \in \mathcal{G}_{\mu, \sigma}^{n \times r}$. So (cf. (2.8))

$$F_{1/\|(MH)+\|}(y) = F_{\sigma_{\hat{r}}(MH)}(y). \quad (4.1)$$

Let $M = S_M \Sigma_M T_M^T$ be full SVD where $\Sigma_M = \text{diag}(\hat{\Sigma}_M, O) = \Sigma_M \text{diag}(I_{r(M)}, O)$ and $\hat{\Sigma}_M = \text{diag}(\sigma_j(M))_{j=1}^{r(M)}$ is a nonsingular diagonal matrix. We have $MH = S_M \Sigma_M T_M^T H$, and so $\sigma_j(MH) = \sigma_j(\Sigma_M T_M^T H)$ for all j by virtue of Lemma 4.1, because S_M is a square orthogonal matrix. Write $H_{r(M)} = (I_{r(M)} \mid O) T_M^T H$ and observe that $\sigma_j(\Sigma_M T_M^T H) = \sigma_j(\hat{\Sigma}_M H_{r(M)})$ and consequently

$$\sigma_j(MH) = \sigma_j(\hat{\Sigma}_M H_{r(M)}) \text{ for all } j. \quad (4.2)$$

Combine equation (4.2) for $j = \hat{r}$ with Lemma 4.2 for the pair (Σ, H) replaced by $(\hat{\Sigma}_M, H_{r(M)})$ and obtain that $\sigma_{\hat{r}}(MH) \geq \sigma_{r(M)}(M)\sigma_{\hat{r}}(H_{r(M)}) = \sigma_{\hat{r}}(H_{r(M)})/\|M^+\|$. We have $T_M^T H' \in \mathcal{G}_{\mu, \sigma}^{n \times r}$ by virtue of Lemma 4.3, because T_M is a square orthogonal matrix; consequently $H_{r(M)} = H'_{r(M)} + B$ for $H'_{r(M)} \in \mathcal{G}_{\mu, \sigma}^{r(M) \times r}$ and some matrix B . Therefore we can apply Theorem 3.1 for $A = H'_{r(M)}$ and obtain the bound of Theorem 4.1 on $F_{1/\|(MH)+\|}(y)$. One can similarly deduce the bound on $F_{1/\|(GM)+\|}(y)$ or can just apply the above bound on $F_{1/\|(MH)+\|}(y)$ for $H = G^T$ and M replaced by M^T and then recall that $(M^T G^T)^T = GM$. \square

By combining (2.4) with Theorems 3.2 (for $B = O$) and 4.1 we can probabilistically bound the condition numbers of randomized products GM and MH . The following corollary extends the bound of Theorem 4.1 for a randomized matrix product to the bounds for its leading blocks.

Corollary 4.2. *Suppose j, k, m, n, q and s are integers, $1 \leq j \leq q, 1 \leq k \leq s, M \in \mathbb{R}^{m \times n}, \sigma > 0, G \in \mathcal{G}_{\mu, \sigma}^{q \times m}, H \in \mathcal{G}_{\mu, \sigma}^{n \times s}, \text{rank}(M_j) = j$ for $M_j = M \begin{pmatrix} I_j \\ O_{n-j, j} \end{pmatrix}, \text{rank}(M^{(k)}) = k$ for $M^{(k)} = (I_k \mid O_{k, m-k})M$, and $y \geq 0$. Then (i) with probability 1 the matrix GM (resp. MH) has generic rank profile if $\text{rank}(M) \geq q$ (resp. if $\text{rank}(M) \geq s$). Furthermore (ii) $F_{1/\|((GM)_j^{(j)})^+\|}(y) \leq 2.35y\sqrt{j}/(\|M_j^+\|\sigma)$ if $\text{rank}(M) \geq j$, $F_{1/\|((MH)_k^{(k)})^+\|}(y) \leq 2.35y\sqrt{k}(\|(M^{(k)})^+\|\sigma)$ if $\text{rank}(M) \geq k$.*

Proof. We immediately verify part (i) by applying the techniques of Section 3.2. To prove part (ii) apply Theorem 4.1 replacing G by $(I_j \mid O_{j, q-j})G$ and replacing M by $M \begin{pmatrix} I_j \\ O_{n-j, j} \end{pmatrix}$. For every k apply Theorem 4.1 replacing M by $(I_k \mid O_{k, m-k})M$ and replacing H by $H \begin{pmatrix} I_k \\ O_{s-k, k} \end{pmatrix}$. \square

Remark 4.1. *It is well known that GENP and block Gaussian elimination are numerically unsafe where the input matrix M has a singular or ill conditioned leading block, but if this matrix itself is well conditioned, then the latter results combined with (2.4) and Theorems 2.5 and 3.2 for $B = O$ imply that multiplication by Gaussian random matrices is expected to fix this problem. Namely both elimination algorithms applied to the matrices GM and MH for $G \in \mathcal{G}_{0,1}^{m \times m}$ and $H \in \mathcal{G}_{0,1}^{n \times n}$ are expected to use no divisions by absolutely small values.*

Remark 4.2. *We cannot extend the proofs of Lemma 4.3 and consequently Theorem 4.1 and its corollaries to the case of Gaussian random Toeplitz matrices $G \in \mathcal{T}_{\mu, \sigma}^{r \times m}$ and $H \in \mathcal{T}_{\mu, \sigma}^{n \times r}$, but the results of our tests have consistently supported such extensions (cf. Tables 10.6 and 10.10). This is also the case for our results in the next two sections.*

5 Randomized additive and dual additive preconditioning

In this section we prove Theorem 1.1 and extend it to the cases of rectangular $m \times n$ matrices A and dual additive preprocessing. At first we prove the following specification of the theorem where instead of the matrix A having numerical rank ρ we deal with its SVD truncation having rank ρ .

Theorem 5.1. *Suppose A is a real $n \times n$ matrix of a rank $\rho, 0 < \rho < n, \sigma U$ and σV are standard Gaussian random $n \times r$ matrices, whose all $2nr$ entries are independent of each other, $0 < r < n$, and $C = A + UV^T$. Then (i) we have $\|A\|_2 - \|U\|_2 \|V\|_2 \leq \|C\|_2 \leq \|A\|_2 + \|U\|_2 \|V\|_2$, which implies (1.2), (ii) the matrix C is singular if $r < n - \text{rank}(A)$, (iii) otherwise it is nonsingular with probability 1, and (iv) the value $\sigma_n(C)$ is expected to have at most order $\sigma_\rho(A)$ if the ratio $\sigma/\|A\|_2$ is neither large nor small, e.g., if $\frac{1}{100} \leq \sigma/\|A\|_2 \leq 100$.*

5.1 Proof of Theorem 5.1 (parts (i)–(iv), case $r = n - \rho$)

Part (i) is an immediate observation, which implies (1.2) by virtue of (3.1) and Theorem 3.2 for $A = U, A = V$ and $h = n$. Furthermore we readily prove parts (ii) and (iii) of the theorem (on singularity and nonsingularity) by applying the techniques of Section 3.2. To prove part (iv), that is to bound the ratio $\|C^{-1}\|/\|A^+\|$, we factorize the matrix C , which involves a number of technicalities. In this subsection we only handle the case where $r = n - \rho$.

Theorem 5.2. *Suppose $A, C, S, T \in \mathbb{R}^{n \times n}$ and $U, V \in \mathbb{R}^{n \times r}$ for two positive integers r and $n, r \leq n, A = S\Sigma T^T$ is full SVD of the matrix A (cf. (2.5)), S and T are square orthogonal matrices, $\Sigma = \text{diag}(\sigma_j)_{j=1}^n$, the matrix $C = A + UV^T$ is nonsingular, and so $\rho = \text{rank}(A) = n - r$ and $\sigma_\rho > 0$. Write*

$$S^T U = \begin{pmatrix} \bar{U} \\ U_r \end{pmatrix}, \quad T^T V = \begin{pmatrix} \bar{V} \\ V_r \end{pmatrix}, \quad R_U = \begin{pmatrix} I_\rho & \bar{U} \\ O_{r, \rho} & U_r \end{pmatrix}, \quad R_V = \begin{pmatrix} I_\rho & \bar{V} \\ O_{r, \rho} & V_r \end{pmatrix}, \quad (5.1)$$

where U_r and V_r are $r \times r$ matrices. Then

(a) $R_U \Sigma R_V^T = \Sigma$, whereas $R_U \text{diag}(O_{\rho, \rho}, I_r) R_V^T = S^T U V^T T$, and so

$$C = S R_U D R_V^T T^T, \quad D = \Sigma + \text{diag}(O_{\rho, \rho}, I_r) = \text{diag}(d_j)_{j=1}^n \quad (5.2)$$

where $d_j = \sigma_j$ for $j = 1, \dots, \rho$, $d_j = \sigma_j + 1$ for $j = \rho + 1, \dots, n$.

Furthermore suppose that the matrix A has been normalized so that $\|A\| = 1$ and that the $r \times r$ matrices U_r and V_r are nonsingular, which holds with probability 1 where U and V are Gaussian random matrices (cf. Section 3.2). Write

$$p = \|R_U^{-1}\| \|R_V^{-1}\| \text{ and } f_r = \max\{1, \|U_r^{-1}\|\} \max\{1, \|V_r^{-1}\|\}. \quad (5.3)$$

Then

- (b) the matrix C is nonsingular,
- (c) $1 \leq \|R_V\| \|R_U\| \leq \sigma_\rho(A)/\sigma_n(C) \leq p$,
- (d) $p \leq (1 + \|U\|)(1 + \|V\|)f_r$,
- (e) $1 \leq \sigma_\rho(A)/\sigma_n(C) \leq (1 + \|U\|)(1 + \|V\|)f_r$.

Proof. Parts (a) and (b) are readily verified.

(c) Combine the equations $S^{-1} = S^T$, $T^{-1} = T$ and (5.2) and obtain $C^{-1} = T R_V^{-T} D^{-1} R_U^{-1} S^T$ or equivalently $D^{-1} = R_V T^T C^{-1} S R_U$. It follows that $\|C^{-1}\| = \|R_V^{-T} D^{-1} R_U^{-1}\|$ and $\|D^{-1}\| = \|R_V T^T C^{-1} S R_U\|$. Apply bound (2.4), substitute $\|S\| = \|S^T\| = \|T\| = \|T^T\| = 1$ and obtain $\|C^{-1}\| \leq \|R_V^{-T}\| \|D^{-1}\| \|R_U^{-1}\|$ and $\|D^{-1}\| \leq \|R_V\| \|C^{-1}\| \|R_U\|$. Substitute the equations (5.3), $\|D^{-1}\| = 1/\sigma_\rho(A)$ (implied by the equations $\|A\| = 1$ and (5.2)) and $\|C^{-1}\| = 1/\sigma_n(C)$ and the bounds $\|R_V\| \geq 1$ and $\|R_U\| \geq 1$ and obtain that $1 \leq \sigma_\rho(A)/\sigma_n(C) \leq p$.

(d) Observe that $R_U^{-1} = \begin{pmatrix} I_\rho & -\bar{U} \\ O & I_r \end{pmatrix} \begin{pmatrix} I_\rho & O \\ O & U_r^{-1} \end{pmatrix}$, $R_V^{-1} = \begin{pmatrix} I_\rho & -\bar{V} \\ O & I_r \end{pmatrix} \begin{pmatrix} I_\rho & O \\ O & V_r^{-1} \end{pmatrix}$, $\|\bar{U}\| \leq \|U\|$ and $\|\bar{V}\| \leq \|V\|$. Then combine these relationships.

(e) Combine the bounds of parts (c) and (d). \square

We have $\frac{\kappa(C)}{\kappa(A)} \leq \frac{\|C\|}{\|A\|} \frac{\sigma_\rho(A)}{\sigma_n(C)}$, and so parts (d) and (e) together bound the ratio $\frac{\kappa(C)}{\kappa(A)}$ in terms of the norms $\|U\|$, $\|V\|$, $\|U_r^{-1}\|$ and $\|V_r^{-1}\|$ as follows,

$$\frac{\kappa(C)}{\kappa(A)} \leq (1 + \|U\| \|V\|)(1 + \|U\|)(1 + \|V\|) \max\{1, \|U_r^{-1}\|\} \max\{1, \|V_r^{-1}\|\}, \quad (5.4)$$

and in particular

$$\frac{\kappa(C)}{\kappa(A)} \leq (1 + \|U\|^2)(1 + \|U\|)^2 \max\{1, \|U_r^{-1}\|^2\} \text{ if } U = V. \quad (5.5)$$

Let us estimate the norms $\|U\|$, $\|V\|$, $\|U_r^{-1}\|$ and $\|V_r^{-1}\|$ where U and V are Gaussian random matrices.

Theorem 5.3. Suppose that A , U , V , U_r and V_r denote the five matrices of Theorem 5.2 where $U, V \in \mathcal{G}_{\mu, \sigma}^{n \times r}$. Then $\max\{F_{1/\|U_r^{-1}\|}(y), F_{1/\|V_r^{-1}\|}(y)\} \leq 2.35 y\sqrt{r}/\sigma$ for $y \geq 0$.

Proof. Lemma 4.3 implies that $S^T U, T^T V \in \mathcal{G}_{\mu, \sigma}^{n \times r}$ by virtue of Lemma 4.3, because S and T are square orthogonal matrices. Hence $U_r, V_r \in \mathcal{G}_{\mu, \sigma}^{r \times r}$. Apply Theorem 3.1 for $A = U_r$ and $A = V_r$ where in both cases $m = n = r$. \square

Combine Theorem 5.3 with relationships (5.3)–(5.5) and obtain part (iv) of Theorem 5.1 in the case where $r = n - \rho$ and $\|A\| = 1$. Relax the normalization assumption by scaling the matrix A .

Remark 5.1. So far our proof of Theorem 5.1 remains valid where $U = V \in \mathcal{G}_{0,1}^{n \times r}$, but not so in the case where $r > n - \rho$, covered in the next subsection. We can make similar comments on the extension to the proofs of Theorems 1.1 and 5.6 in Section 5.4.

5.2 Proof of Theorem 5.1 (part (iv), case $r > n - \rho$)

Next we extend the proof of part (iv) to the case where $s = r + \rho - n > 0$. The extension is immediate where the matrix A is nonnegative definite and $U = V$ but is more involved and less transparent in the general case. Write $U = (U^{(s)} \mid U_s)$ and $V = (V^{(s)} \mid V_s)$ where $U^{(s)}, V^{(s)} \in \mathcal{G}_{0,\sigma}^{(n-s) \times r}$ and $U_s, V_s \in \mathcal{G}_{0,\sigma}^{s \times r}$. As we have proved already, the $n \times n$ matrices $C^{(s)} = A + U^{(s)}V^{(s)T}$ and $C = A + UV^T = C^{(s)} + U_sV_s^T$ are nonsingular with probability 1 and are expected to have norms of order $\|A\|$ for reasonably bounded values σ (cf. (1.2)), whereas the matrix $C^{(s)}$ is expected to be well conditioned, that is the ratio $\kappa_s = \|C^{(s)}\|/\sigma_n(C^{(s)}) \geq 1$ is not expected to be large. To simplify the notation, scale the matrices A, C, U and V to have $\sigma = 1$, expecting that the scaling factor and the new value of the norm $\|C^{(s)}\|$ is neither large nor small (cf. (3.1)).

Let $C^{(s)} = S\Sigma T^T$ be SVD, where $\Sigma = \text{diag}(\sigma_j(C^{(s)}))_{j=1}^n$, premultiply the equation $C = C^{(s)} + U_sV_s^T$ by S^T , postmultiply it by T , write $\hat{C} = S^TCT$, $\hat{U} = S^TU_s$, and $\hat{V}^T = V_s^TT$, and obtain $\hat{C} = \Sigma + \hat{U}\hat{V}^T$ where $\sigma_j(\hat{C}) = \sigma_j(C)$ for all j by virtue of Lemma 4.1 and $\hat{U}, \hat{V} \in \mathcal{G}_{0,1}^{n \times r}$ by virtue of Lemma 4.3, because S and T are square orthogonal matrices. Furthermore write $\delta_{ij} = 0$ for $i \neq j$, $\delta_{ii} = 1$, $\mathbf{c}_i = \hat{C}^T \mathbf{e}_i$, $\mathbf{c}_j^- = \hat{C}^{-1} \mathbf{e}_j$, for $i, j = 1, \dots, n$, and so

$$\mathbf{c}_i^T \mathbf{c}_j^- = \mathbf{e}_i^T \mathbf{e}_j = \delta_{ij} \text{ for } i, j = 1, \dots, n, \quad (5.6)$$

$$\|\mathbf{c}_j^-\| = 1/\mathbf{c}_j^T \hat{\mathbf{c}}_j \text{ for } j = 1, \dots, n. \quad (5.7)$$

Indeed for every j the unit vector $\hat{\mathbf{c}}_j = \mathbf{c}_j^-/\|\mathbf{c}_j^-\| = (\hat{c}_{ij})_{i=1}^n$ is unique because the vector \mathbf{c}_j^- is orthogonal to all vectors \mathbf{c}_i for $i \neq j$ (cf. (5.6)). Combine the latter equation $\hat{\mathbf{c}}_j = \mathbf{c}_j^-/\|\mathbf{c}_j^-\|$ with $\mathbf{c}_j^T \mathbf{c}_j^- = 1$ and deduce equation (5.7) (cf. [SST06, Proof of Lemma 3.2]).

Next we estimate the norm $\|\mathbf{c}_j^-\|$ from above or equivalently the value $\mathbf{c}_j^T \hat{\mathbf{c}}_j$ from below for any fixed integer j , $1 \leq j \leq n$. We write $\hat{\mathbf{u}}_j^T = \mathbf{e}_j^T \hat{U} \in \mathcal{G}_{0,\sigma}^{n \times 1}$ and $\hat{\mathbf{t}}_j = \hat{V}^T \hat{\mathbf{c}}_j$ and are going to deduce a probabilistic upper bound on the norm $\|\hat{\mathbf{t}}_j\|$ as long as we have a probabilistic upper bound on the norm $\|\mathbf{c}_j^-\|$. Represent the value $\mathbf{c}_j^T \hat{\mathbf{c}}_j$ as $\mathbf{e}_j^T \hat{C} \hat{\mathbf{c}}_j = \mathbf{e}_j^T \Sigma \hat{\mathbf{c}}_j + \mathbf{e}_j^T \hat{U} \hat{V}^T \hat{\mathbf{c}}_j$ and infer that

$$\mathbf{c}_j^T \hat{\mathbf{c}}_j = \sigma_j(C^{(s)}) \hat{c}_{jj} + \hat{\mathbf{u}}_j^T \hat{\mathbf{t}}_j. \quad (5.8)$$

Recall Lemma 3.2 for $\sigma = 1$ and $\mathbf{b} = \mathbf{u}_j$, recall Remark 3.1, and obtain that $F_{\mathbf{c}_j^T \hat{\mathbf{c}}_j}(y) \leq \sqrt{\frac{2}{\pi}} \frac{y}{\|\hat{\mathbf{t}}_j\|}$. Write $p = F_{\mathbf{c}_j^T \hat{\mathbf{c}}_j}(y)$ and infer that

$$\|\hat{\mathbf{t}}_j\| \leq \sqrt{\frac{2}{\pi}} \frac{y}{p}. \quad (5.9)$$

Next deduce upper estimates for the values $|\hat{c}_{ij}|$ for all i , at first for $i = j$. Obtain from equation (5.8) that $\hat{c}_{jj} = (\mathbf{c}_j^T \hat{\mathbf{c}}_j - \hat{\mathbf{u}}_j^T \hat{\mathbf{t}}_j)/\sigma_j(C^{(s)})$. Substitute $\mathbf{c}_j^T \hat{\mathbf{c}}_j \leq y$ and (5.9) and obtain $|\hat{c}_{jj}| \leq (y + \|\hat{\mathbf{u}}_j\| \|\hat{\mathbf{t}}_j\|)/\sigma_j(C^{(s)}) \leq (y + \|\hat{\mathbf{u}}_j\| \sqrt{\frac{2}{\pi}} \frac{y}{p})/\sigma_j(C^{(s)})$, and so

$$|\hat{c}_{jj}| \leq (1 + \sqrt{\frac{2}{\pi}} \frac{\|\hat{\mathbf{u}}_j\|}{p}) y \kappa_s, \quad (5.10)$$

where the value $\kappa_s = 1/\sigma_n(C^{(s)})$ is not expected to be large and where the cdf $F_{\|\hat{\mathbf{u}}_j\|}(y) = \chi_{0,1,n}(y)$ is bounded in Definition 3.3.

Next let $i \neq j$, recall that $\mathbf{c}_i^T \mathbf{c}_j^- = 0$ (cf. (5.6)), substitute $\mathbf{c}_i^T = \mathbf{e}_i^T \hat{C} = \mathbf{e}_i^T \Sigma + \hat{\mathbf{u}}_i^T \hat{V}^T$ and obtain $\mathbf{c}_i^T \mathbf{c}_j^- = \mathbf{e}_i^T \Sigma \mathbf{c}_j^- + \hat{\mathbf{u}}_i^T \hat{V}^T \mathbf{c}_j^- = \sigma_i(C^{(s)}) \hat{c}_{ij} + \hat{\mathbf{u}}_i^T \hat{\mathbf{t}}_j = 0$. Hence $|\hat{c}_{ij}| \leq |\hat{\mathbf{u}}_i^T \hat{\mathbf{t}}_j|/\sigma_i(C^{(s)})$. Substitute (5.9) and obtain

$$|\hat{c}_{ij}| \leq \sqrt{\frac{2}{\pi}} \frac{\|\hat{\mathbf{u}}_i\|}{p} y \kappa_s \text{ for all } i \neq j. \quad (5.11)$$

Combine equations (5.10) and (5.11) and obtain that $\|\hat{\mathbf{c}}_j\|^2 = \sum_{i=1}^n \hat{c}_{ij}^2 \leq \gamma y^2 \kappa_s^2$ where $\gamma = 1 + 2\sqrt{\frac{2}{\pi}} \frac{\|\hat{\mathbf{u}}_j\|}{p} + \frac{2}{\pi} \sum_{i=1}^n \|\hat{\mathbf{u}}_i\|^2/p^2$ and $F_{\|\hat{\mathbf{u}}_i\|}(y) = \chi_{0,1,n}(y)$ for all i . Recall that $\hat{\mathbf{c}}_j$ is a unit

vector, and consequently $\gamma y^2 \kappa_s^2 \geq 1$, which probabilistically bounds the ratio y/p from below, thus implying the desired upper bounds on the norms $\|\mathbf{c}_j^-\| = \|\hat{C}^{-1} \mathbf{e}_j\|$ for all j and consequently $\|C^{-1}\| = \|\hat{C}^{-1}\| \leq \sum_{j=1}^n \|\hat{C}^{-1} \mathbf{e}_j\|$. This completes our proof of part (iv) of Theorem 5.1.

5.3 Extension of Theorem 5.1 to the case of rectangular matrices

Clearly part (i) of Theorem 5.1 holds for any pair of m and n . By extending the concept of singularity of a matrix to its rank deficiency, we readily extend parts (ii) and (iii). Next we employ Fact 2.3 and Lemma 4.3 to extend our upper bound on $\sigma_\rho(A)/\sigma_n(C)$ to the case where $m \neq n$.

Theorem 5.4. *Suppose $A \in \mathbb{R}^{m \times n}$, $U \in \mathcal{G}_{0,\sigma}^{m \times r}$, and $V \in \mathcal{G}_{0,\sigma}^{n \times r}$ for three positive integers m , n and r , the matrix $C = A + UV^T$ has full rank $l = \min\{m, n\}$, and $l - r \leq \rho = \text{rank}(A)$. Keep equation (5.1) but write*

$$I_{m,n} S^T U = \begin{pmatrix} \bar{U} \\ U_r \end{pmatrix}, \quad I_{m,n} T^T V = \begin{pmatrix} \bar{V} \\ V_r \end{pmatrix} \quad (5.12)$$

for $I_{g,h}$ of (2.1) where U_r and V_r still denote $r \times r$ matrices. Keep the other assumptions of parts (a)–(e) of Theorem 5.2. Then the upper bound of part (e) of Theorem 5.2 can be extended, that is, $\sigma_\rho(A)/\sigma_l(C) \leq (1 + \|U\|)(1 + \|V\|)f_r$ where $f_r = \max\{1, \|\bar{U}_r^{-1}\|\} \max\{1, \|\bar{V}_r^{-1}\|\}$ as in (5.3) and where $U_r, V_r \in \mathcal{G}_{0,\sigma}^{r \times r}$.

Proof. Let $A = S_A \Sigma_A T_A^T$ be SVD of (2.5). Write $\hat{C} = I_{m,n} S_A^T C T_A I_{n,m}^T$, $\hat{U} = I_{m,n} S_A^T U$, $\hat{V} = I_{n,m} T_A^T V$, $\hat{A} = I_{m,n} S_A^T A T_A I_{n,m}^T$, and so $\hat{A} = (\sigma_j(A))_{j=1}^l$ and $\hat{C} = \hat{A} + \hat{U} \hat{V}^T$. Apply Theorem 5.2 to the $l \times l$ matrices \hat{A} and \hat{C} and obtain that $\sigma_\rho(\hat{A})/\sigma_n(\hat{C}) \leq (1 + \|\hat{U}\|)(1 + \|\hat{V}\|)f_r$. Complete the proof of Theorem 5.4 by combining this bound with the relationships $\sigma_\rho(\hat{A}) = \sigma_\rho(A)$, $\sigma_l(\hat{C}) \leq \sigma_l(S_A^T C T_A) = \sigma_l(C)$, $\|U\| = \sigma_1(U) \geq \sigma_1(\hat{U}) = \|\hat{U}\|$, and $\|V\| = \sigma_1(V) \geq \sigma_1(\hat{V}) = \|\hat{V}\|$. Here the equations hold by virtue of Lemma 4.1, because the matrices S_A and T_A are square and orthogonal. The inequalities hold by virtue of Fact 2.3, because \hat{C} , \hat{U} , and \hat{V} are submatrices of the matrices $S_A^T C T_A$, $S_A^T U$, and $V^T T_A$, respectively. \square

Combine Theorems 5.3 and 5.4 to yield the following result.

Theorem 5.5. *Assume that $A \in \mathbb{R}^{m \times n}$, $U \in \mathcal{G}_{0,\sigma}^{m \times r}$, $V \in \mathcal{G}_{0,\sigma}^{n \times r}$, $C = A + UV^T$, and $l = \min\{m, n\}$. Then the matrix C is rank deficient if $r < l - \text{rank}(A)$. Otherwise with probability 1 the matrices U and V have full rank and the bound $\sigma_\rho(A)/\sigma_l(C) \leq (1 + \|U\|)(1 + \|V\|)f_r$ of Theorem 5.4 holds, the norms $\|U\|$ and $\|V\|$ satisfy the randomized bounds of Theorem 3.2 (for $A = U$ and $A = V$), and the values $\|U_r^{-1}\|$ and $\|V_r^{-1}\|$ satisfy the randomized bounds of Theorem 5.3.*

Corollary 5.1. *Theorem 5.1 can be extended to the case of matrices $A \in \mathbb{R}^{m \times n}$, $U \in \mathcal{G}_{0,\sigma}^{m \times r}$ and $V \in \mathcal{G}_{0,\sigma}^{n \times r}$ for any pair of positive integers $\{m, n\}$ and $l = \min\{m, n\}$, that is the matrix $C = A + UV^T$ is rank deficient if $r < l - \text{rank}(A)$, whereas for $r \geq l - \text{rank}(A)$ it has full rank with probability 1 and is expected to have condition number of at most order $\|A\|/\sigma_{l-r}(A)$ if the ratio $\sigma/\|A\|_2$ is neither large nor small, e.g., if $\frac{1}{100} \leq \sigma/\|A\|_2 \leq 100$. Consequently the matrix C is expected to be nonsingular and well conditioned if the matrix A has numerical rank at least $l - r$.*

5.4 Extension to and from Theorem 1.1

To extend Theorem 5.1 to Theorem 1.1 truncate the SVD of the matrix A having numerical rank $\rho < n$ by setting to 0 all its singular values, except for the ρ largest ones. This produces a well conditioned matrix $A - E$ of rank ρ where $\|E\| = \sigma_{\rho+1}(A)$ and the ratio $\|E\|/\|A\|$ is small because the matrix A has numerical rank ρ . Now to obtain Theorem 1.1 combine Theorem 5.1 (applied to the matrix $A - E$ rather than A) with Theorem 2.3 and the simple bounds $\|A\| - \|E\| \leq \|A\| \leq \|A\| + \|E\|$ and $\|C\| - \|E\| \leq \|C\| \leq \|C\| + \|E\|$ and observe that an upper bound of at most order $\|A\|/\sigma_\rho(A)$ on $\kappa(A)$ implies that the matrix A having numerical rank ρ is well conditioned.

We can apply Theorem 2.2 instead of Theorem 2.3 and similarly extend the results of the previous subsection, to rectangular matrices A . To yield stronger estimates, however, one should avoid using Theorem 2.2 and instead extend Theorem 1.1 to the case of rectangular input by applying our techniques of the proof of Theorem 5.4. Here is the resulting extension of Theorem 1.1.

Theorem 5.6. *Suppose A is a real $m \times n$ matrix having a numerical rank ρ (that is the ratio $\sigma_{\rho+1}(A)/\|A\|$ is small, but the ratio $\sigma_\rho(A)/\|A\|$ is not small), the $(m+n)r$ Gaussian random entries of two matrices $U \in \mathcal{G}_{0,\sigma}^{m \times r}$ and $V \in \mathcal{G}_{0,\sigma}^{n \times r}$, are independent of each other, $C = A + UV^T$, $0 < r < l$ and $0 < \rho < l = \min\{m, n\}$. Then (i) bounds of (1.2) hold, (ii) the matrix C is singular or ill conditioned if $r < l - \rho$; (iii) otherwise it is nonsingular with probability 1 and (iv) is expected to be well conditioned if the ratio $\sigma/\|A\|_2$ is neither large nor small, e.g., if $\frac{1}{100} \leq \sigma/\|A\|_2 \leq 100$.*

The next corollary slightly generalizes Theorem 5.6 to match the augmentation map of Theorem 6.2 of the next section.

Corollary 5.2. *Suppose that A , U and V denote the same matrices as in Theorem 5.6, l still denotes the integer $\min\{m, n\}$, but $C = A + UMV^T$ for a normalized nonsingular $r \times r$ matrix M , $\|M\| = 1$. Then Theorem 5.6 can be extended as follows: the matrix C is rank deficient if $r < l - \rho$; otherwise it has full rank with probability 1 and is expected to have condition number of order $\kappa(M)$ if the ratio $\sigma/\|A\|_2$ is neither large nor small, e.g., if, say $\frac{1}{100} \leq \sigma/\|A\|_2 \leq 100$.*

Proof. Let $M = \Sigma_M T_M^T$ be SVD and rewrite $C = A + UMV^T$ as $C = A + \bar{U}\bar{V}^T$ where $\bar{U} = US_M$ and $\bar{V} = \Sigma_M T_M^T V$. Note that $\bar{U} \in \mathcal{G}_{0,\sigma}^{m \times r}$ and $T_M V \in \mathcal{G}_{0,\sigma}^{n \times r}$ by virtue of Lemma 4.3. Now reapply the proofs of this section replacing U by \bar{U} and V by \bar{V} . All the proofs are readily extended except for the estimates for the norm V_r^{-1} , which grow by at most a factor $\kappa(\Sigma_M) = \kappa(M)$ by virtue of Lemma 4.2. \square

Remark 5.2. *How large is the class of $m \times n$ matrices having a numerical rank ρ ? We characterize it indirectly, by noting that by virtue of Fact 2.2 the nearby matrices of rank ρ form a variety of dimension $(m+n-\rho)\rho$, which increases as ρ increases.*

5.5 Dual additive preconditioning

For an $m \times n$ matrix A of full rank we extend (2.9) and (2.10) to define the *dual additive preprocessing*

$$A^+ \implies C_-^+ = A^+ + U_- V_-^T. \quad (5.13)$$

Our analysis implies that the value $\kappa(C_-^+)$ (equal to $\kappa(C_-)$) is expected to have at most order $\sigma_{q+1}(A)/\sigma_l(A)$ provided $l = \min\{m, n\}$, $U_- \in \mathcal{G}_{0,1}^{n \times q}$, $V_- \in \mathcal{G}_{0,1}^{m \times q}$, and the norm $\|A^+\|$ is neither large nor small. The randomized algorithm of [D83] is expected to estimate the norm $\|A^+\|$ at a low computational cost. We can work with the $(m+1) \times (n+1)$ matrix $\hat{A} = \text{diag}(A, \epsilon)$ instead of the matrix A and choose a sufficiently small positive scalar ϵ such that $\|\hat{A}^+\| = 1/\epsilon$. Then we can scale the matrix \hat{A} to obtain that $\|(\hat{A}/\epsilon)^+\| = 1$.

5.6 Can we weaken randomness?

Would Theorems 1.1, 5.1, 5.6 and other results of this section and of the next one still hold if we weaken randomness of the matrices U and V by allowing them to be sparse and structured, to share some or all their entries, or generally to be defined by a smaller number of independent parameters, possibly under other probability distributions rather than Gaussian? We have some progress with our analytical study in this direction (see Sections 3.4 and 3.5 and Remark 5.1), but empirically all the presented randomized techniques remain as efficient under very weak randomization in the above sense (cf. Tables 10.5, 10.6, 10.10, and 10.16).

6 Randomized augmentation

6.1 Augmentation and an extension of the SMW formula

The solution of a nonsingular linear system of n equations, $A\mathbf{y} = \mathbf{b}$ can be readily recovered from a null vector $\begin{pmatrix} -1/\beta \\ \mathbf{y} \end{pmatrix}$ of the matrix $K = (\beta\mathbf{b} \mid A)$ for a nonzero scalar β . If the matrix A has numerical nullity 1 and if the ratio $\|A\|/\|\beta\mathbf{b}\|$ is neither large nor small, then the matrix K is well conditioned for the average vector \mathbf{b} [PQa, Section 13.1]. The above map $A \Rightarrow K$ is a special case of more general augmentation

$$K = \begin{pmatrix} W & V^T \\ -U & A \end{pmatrix}, \quad (6.1)$$

which we study next, beginning with the following extension of the SMW formula.

Theorem 6.1. *Suppose equation (6.1) holds, $m = n$ and the matrices A , W and K are nonsingular. Write $S = A + UW^{-1}V^T$ and $R = I - V^T S^{-1}UW^{-1}$. Then the matrix S is nonsingular, S^{-1} is the trailing (southwestern) $n \times n$ block of K^{-1} , and*

$$A^{-1} = S^{-1} + S^{-1}UW^{-1}R^{-1}V^T S^{-1}. \quad (6.2)$$

Proof. Apply the SMW formula of Theorem 2.4 for C replaced by S , U by UW^{-1} , and G by R . \square

6.2 Links to additive preprocessing and condition estimates

In contrast to the scaled randomized symmetric additive preprocessing $A \Rightarrow C = A + VV^T$ (cf. (5.5) and [W07]), the map $A \Rightarrow K = \begin{pmatrix} W & V^T \\ V & A \end{pmatrix}$ cannot decrease the condition number $\kappa(A)$ if K is a symmetric and positive definite matrix; this follows from the Interlacing Property of the eigenvalues of K [GL96, Theorem 8.6.3]. Nonetheless the following simple theorem links additive preprocessing $A \Rightarrow C = A + UMV^T$ to the augmentation $A \Rightarrow K$ for K of (6.1) and later we extend Theorem 5.6 to the augmentation as well.

Theorem 6.2. *Suppose $A \in \mathbb{R}^{m \times n}$, $W \in \mathbb{R}^{r \times r}$, the matrix W is nonsingular, $l = \min\{m, n\}$, a matrix K in $\mathbb{R}^{(m+r) \times (n+r)}$ is defined by (6.1), and $C = A + UW^{-1}V^T$. Then we have*

$$K = \widehat{U} \operatorname{diag}(C, I_r) \widehat{V} \operatorname{diag}(W, I_n) \quad (6.3)$$

for $\widehat{U} = \begin{pmatrix} O_{r,m} & I_r \\ I_m & -UW^{-1} \end{pmatrix}$, $\widehat{V} = \begin{pmatrix} O_{n,r} & I_n \\ I_r & V^T \end{pmatrix}$, the matrix C has full rank if and only if the matrix K has full rank, and so both matrices are rank deficient for $r < l$. Furthermore $\widehat{U}^{-1} = \begin{pmatrix} UW^{-1} & I_n \\ I_r & O_{r,n} \end{pmatrix}$, $\widehat{V}^{-1} = \begin{pmatrix} -V^T & I_r \\ I_n & O_{n,r} \end{pmatrix}$. For $m = n$ and nonsingular matrices C and K , we have $C^{-1} = (I_n \mid O_{n,r}) \widehat{V} \operatorname{diag}(W, I_n) K^{-1} \widehat{U} (I_n \mid O_{n,r})^T$ and $K^{-1} = \operatorname{diag}(W^{-1}, I_n) \widehat{V}^{-1} \operatorname{diag}(C^{-1}, I_r) \widehat{U}^{-1}$.

Corollary 6.1. (Cf. [PQa, Remark 10.1 and Corollary 11.1].) *Define three integers m , n , and l and three matrices A , K , and W as in Theorem 6.2, write $h = \max\{m, n\}$, and suppose that $U \in \mathcal{G}_{0,\sigma}^{m \times r}$ and $V \in \mathcal{G}_{0,\sigma}^{n \times r}$. Then (i) $\|A\| \leq \|K\| \leq \|A\| + \|U\| + \|V\| + \|W\|$, and (ii) the matrix K is rank deficient if $r < l - \operatorname{rank}(A)$ but has full rank with probability 1 otherwise. (iii) Furthermore suppose $r \geq l - \operatorname{rank}(A)$ and the ratio $\sigma/\|A\|_2$ is neither large nor small, e.g., say $\frac{1}{100} \leq \sigma/\|A\|_2 \leq 100$. Then the matrix K is expected to have condition number of order $(1 + 2h)^4 \kappa(W)^2 / \sigma_{l-r}(A)$, that is (iv) of order $(1 + 2h\sigma)^4 / \sigma_{l-r}(A)$ provided $W \in \mathcal{G}_{0,1}^{r \times r}$.*

Proof. Part (i) is verified immediately. Next estimate the rank of the matrix C as in parts (ii) and (iii) of Theorem 1.1 and apply equation (6.3) to extend the estimates to the rank of the matrix K . This proves part (ii). Equation (6.3) and Corollary 4.1 together imply that $\kappa(K) \leq \kappa(\widehat{U})\kappa(\widehat{V})\kappa(C)\kappa(W)$. (We can apply Corollary 4.1 because the matrices \widehat{U} , \widehat{V} and W are nonsingular.) We have $\kappa(\widehat{U}) \leq$

$(1 + \|\widehat{U}\|)^2$ and $\kappa(\widehat{V}) \leq (1 + \|\widehat{V}\|)^2$, and we can expect that $\max\{\|\widehat{U}\|, \|\widehat{V}\|\} \leq 1 + 2h\sigma$ for $h = \max\{m, n\}$ by virtue of Theorem 3.2. Now apply Corollary 5.2 for $M = W^{-1}$ to bound $\kappa(C)$ and recall that $\kappa(W^{-1}) = \kappa(W)$. Combining these estimates proves part (iii). To extend part (iii) to part (iv) note that a matrix W in $\mathcal{G}_{0,1}^{r \times r}$ is nonsingular with probability 1 and is expected to be well conditioned (see Sections 3.2 and 3.3). \square

6.3 Direct condition estimation: Gaussian random leading blocks

To obtain sharper bounds and better insight into the subject, let us estimate the condition number $\kappa(K)$ directly, without reducing this task to additive preprocessing. Some initial study of randomized augmentation in this direction can be found in [PQa]. In particular the results of [PQa, Corollary 11.1] are similar to Theorem 6.3, but [PQa] only provides a pointer to the idea of a proof. Part (i) of Corollary 6.1 is extended immediately, and next we extend the other parts.

Theorem 6.3. *Suppose a real normalized $m \times n$ matrix A has a rank $\rho < n$, $U \in \mathcal{G}_{0,1}^{m \times q}$, $V \in \mathcal{G}_{0,1}^{n \times s}$, $W \in \mathcal{G}_{0,1}^{s \times q}$, K in $\mathbb{R}^{(m+s) \times (n+q)}$ defined by (6.1), $l = \min\{m, n\}$, $r = \min\{m - q, n - s\} > 0$. Then*

- (i) *the matrix K is rank deficient if $\rho < r$,*
- (ii) *otherwise the matrix has full rank $l' = \min\{m + s, n + q\}$ with probability 1 and*
- (iii) *is expected to have the condition number $\kappa(K)$ of order at most $1/\sigma_r(A)$.*

Thus we can expect that the matrix K has full rank and is well conditioned if $\rho \geq r$.

Proof. Assume that the entries of the matrices U , V , and W are indeterminates. Then clearly the matrix $(-U \mid A)$ has full rank, that is has m linearly independent rows, if and only if $\rho + q \geq m$.

Likewise the matrix $\begin{pmatrix} V^T \\ A \end{pmatrix}$ has full rank, that is has n linearly independent columns, if and only if $\rho + s \geq n$. The transition from these matrices to the matrix K increases the numbers of linearly independent rows by s and columns by q . Summarizing we obtain parts (i) and (ii) provided that the entries of the matrices U , V , and W are indeterminates. Relax this assumption by applying Lemma 3.1.

Next assume that $\rho + s + q \geq l' = \text{rank}(K)$ and estimate the condition number $\kappa(K) = \|K'\|/\sigma_{l'}(K)$. By virtue of Theorem 3.2 we can expect that the norms of the matrices U , V , and W are in $O(1)$, that is do not exceed a fixed constant, and so $\|K\| = O(1)$ as well because $\|K\| \leq \|U\| + \|V\| + \|W\| + \|A\|$ and $\|A\| = 1$. It remains to estimate the value $\sigma_{l'}(K)$ from below. We can assume that $l' = m + s \leq n + q$, and so $r = m - q$, for otherwise we can estimate $\kappa(K^T) = \kappa(K)$.

At first let $s = 0$. Then $l' = \text{rank}(K) = m \leq n + q$, $K = (-U \mid A)$, and V and W are empty matrices. Reuse and extend the idea of Section 5.3, that is reduce the original task to the case of an $m \times m$ submatrix \bar{K} of the matrix K , which is nonsingular with probability 1 and for which we have $\sigma_m(K) = \sigma_m(\bar{K})$; then estimate the value $\sigma_m(\bar{K})$ as the reciprocal $1/\|\bar{K}^{-1}\|$. Namely assume the SVD $A = S_A \Sigma_A T_A^T$ of (2.5) and write $K' = S_A^T K \text{diag}(I_q, T_A) = (U' \mid \Sigma_A)$. Note that S_A , T_A and $\text{diag}(T_A, I_q)$ are square orthogonal matrices and infer that $\sigma_{l'}(K) = \sigma_m(K')$ by virtue of Lemma 4.1, whereas $U' = S_A^T U \in \mathcal{G}_{0,1}^{m \times q}$ by virtue of Lemma 4.3. The $m \times (n + q)$ matrix K' has the $m \times m$ leading submatrix $\bar{K} = \begin{pmatrix} U_0 & \Sigma_{m-q} \\ U_1 & O_{q, m-q} \end{pmatrix}$ where $U_0 \in \mathcal{G}_{0,1}^{(m-q) \times q}$, $U_1 \in \mathcal{G}_{0,1}^{q \times q}$, $\text{rank}(\bar{K}) = \text{rank}(K) = m$, $\Sigma_{m-q} = \text{diag}(\sigma_j(A))_{j=1}^{m-q}$, and so $\text{rank}(\Sigma_{m-q}) = m - q$ and $\sigma_{l'}(K) = \sigma_m(K) \geq \sigma_m(\bar{K}) = 1/\|\bar{K}^{-1}\|$. We have

$$\bar{K}^{-1} = \begin{pmatrix} O_{q, m-q} & U_1^{-1} \\ \Sigma_{m-q}^{-1} & -\Sigma_{m-q}^{-1} U_0 U_1^{-1} \end{pmatrix} = \text{diag}(I_q, \Sigma_{m-q}^{-1}) \begin{pmatrix} O_{q, n} & I_q \\ I_{m-q} & -U_0 \end{pmatrix} \text{diag}(I_{m-q}, U_1^{-1}).$$

Therefore $\|\bar{K}^{-1}\| \leq \|\Sigma_{m-q}^{-1}\| (1 + \|U_0\|) \|U_1^{-1}\|$ where $\|\Sigma_{m-q}^{-1}\| = 1/\sigma_{m-q}(A)$. Theorems 3.1 and 3.2 together bound the norms $\|U_1^{-1}\|$ and $\|U_0\|$, implying that the value $1/\sigma_{l'}(K) = \|\bar{K}^{-1}\|$ is expected to have at most order $1/\sigma_{m-q}(A)$.

Now let $s > 0$. Then again we reduce our task to the case of a square matrix $\hat{K} \in \mathbb{R}^{l' \times l'}$ (where $l' = m + s$) such that $\sigma_j(K) \geq \sigma_j(\hat{K})$ for all j and then estimate the value $\sigma_{l'}(\hat{K})$ as the reciprocal $1/\|\hat{K}^{-1}\|$. Namely represent the matrix K as $\begin{pmatrix} B \\ F \end{pmatrix}$ where $B = (W \mid V^T)$, $F = (-U \mid A)$, and the value $\|F^+\|$ has at most order $1/\sigma_{m-q}(A)$, as we proved above. Let $F = S_F \Sigma_F T_F^T$ be SVD and write $K'' = \text{diag}(I_s, S_F^T) K T_F = \begin{pmatrix} B_0 & B_1 \\ \hat{\Sigma}_F & O_{m, n+q-m} \end{pmatrix}$ where $B_0 \in \mathcal{G}_{0,1}^{s \times m}$, $B_1 \in \mathcal{G}_{0,1}^{s \times (n+q-m)}$, $\hat{\Sigma}_F = \text{diag}(\sigma_j(F))_{j=1}^m$, $\text{rank}(K'') = \text{rank}(K) = l'$, and so the matrix $\hat{\Sigma}_F$ is nonsingular and $\|\hat{\Sigma}_F^{-1}\| = \|F^+\|$. We have $\sigma_j(K'') = \sigma_j(K)$ for all j because the matrices $\text{diag}(I_s, S_F^T)$ and T_K are square and orthogonal.

Delete the last $n + q - m - s$ columns of the matrix K'' and obtain the $l' \times l'$ submatrix $\hat{K} = \begin{pmatrix} B_0 & \bar{B}_1 \\ \hat{\Sigma}_F & O_{m, n+q-m} \end{pmatrix}$. We have $\sigma_{l'}(K) = \sigma_{l'}(K'') \geq \sigma_{l'}(\hat{K})$. The Gaussian random $s \times s$ matrix \bar{B}_1 is nonsingular with probability 1. We assume that it is nonsingular, and then so is the matrix \hat{K} as well, and consequently $\sigma_{l'}(K) = \sigma_{l'}(\hat{K}) = 1/\|\hat{K}^{-1}\|$. Observe that

$$\hat{K}^{-1} = \begin{pmatrix} O_{m, n+q-m} & \hat{\Sigma}_F^{-1} \\ \bar{B}_1^{-1} & -\bar{B}_1^{-1} B_0 \hat{\Sigma}_F^{-1} \end{pmatrix} = \text{diag}(I_q, \bar{B}_1^{-1}) \begin{pmatrix} O_{q, n} & I_q \\ I_{m-q} & -B_0 \end{pmatrix} \text{diag}(I_{m-q}, \Sigma_F^{-1}).$$

Therefore $\|\hat{K}^{-1}\| \leq \|\bar{B}_1^{-1}\|(1 + \|B_0\|)\|\hat{\Sigma}_F^{-1}\|$. Theorems 3.1 and 3.2 together bound the norms $\|\bar{B}_1^{-1}\|$ and $\|B_0\|$. To complete the proof of the theorem recall that $\|\hat{\Sigma}_F^{-1}\| = \|F^+\|$ and that the norm $\|F^+\|$ is expected to have at most order $1/\sigma_{m-q}(A)$. \square

Compared to Corollary 6.1, Theorem 6.3 allows rectangular matrices W in $\mathcal{G}^{s \times q}$. Combined with Theorem 6.2 it implies Corollary 5.2 restricted to the case where $M^{-1} \in \mathcal{G}_{0,1}^{s \times s}$. In the next subsection we extend Theorem 6.3 by relaxing this restriction provided that $s = q$ and allowing any well conditioned block W with the norm not exceeding 1.

6.4 Direct condition estimates: well conditioned leading blocks

Next we outline a direct proof of Corollary 6.1 allowing any scaled well conditioned square leading blocks W . The supporting estimates are stronger than the ones deduced via combining Corollary 5.2 and Theorem 6.2. We begin with providing all details in the case where $m = n$ and $r = l - \text{rank}(A)$. In this case we allow ill conditioned blocks W .

Theorem 6.4. *Suppose n and r are two positive integers, a real normalized $n \times n$ matrix A has a rank $\rho < n$, $U, V \in \mathcal{G}_{0,1}^{n \times r}$, $W \in \mathbb{R}^{r \times r}$, $\|W\| \leq 1$, and K denotes the matrix of (6.1). Then (i) the matrix K is singular or ill conditioned if $r < n - \rho$. Otherwise it is nonsingular with probability 1. Furthermore (ii) if $r = n - \rho$, then the condition number $\kappa(K)$ is expected to have at most order $\|A\|/\sigma_{n-r}(A)$.*

Proof. We proceed similarly to the proof of Theorem 6.3. Suppose $A = S_A \Sigma_A T_A^T$ is the SVD of (2.5) and write $\bar{K} = \text{diag}(I_r, S_A^T) K \text{diag}(I_r, T_A)$. Then

$$\bar{K} = \begin{pmatrix} W & \bar{V}^T \\ -\bar{U} & \Sigma_A \end{pmatrix}$$

where $\sigma_j(K) = \sigma_j(\bar{K})$ for all j and $\bar{U}, \bar{V} \in \mathcal{G}_{0,1}^{n \times r}$. Furthermore write $\Sigma_\rho = \text{diag}(\sigma_j(A))_{j=1}^\rho$, $\Sigma_A = \text{diag}(\Sigma_\rho, O_{n-\rho, n-\rho})$, $\bar{U} = \begin{pmatrix} U_0 \\ U_1 \end{pmatrix}$ and $\bar{V} = \begin{pmatrix} V_0 \\ V_1 \end{pmatrix}$ where $U_0, V_0 \in \mathcal{G}_{0,1}^{\rho \times r}$ and $U_1, V_1 \in \mathcal{G}_{0,1}^{(n-\rho) \times r}$ and obtain

$$\bar{K} = \begin{pmatrix} W & V_0^T & V_1^T \\ U_0 & \Sigma_\rho & O_{\rho, n-\rho} \\ U_1 & O_{\rho, n-\rho} & O_{n-\rho, n-\rho} \end{pmatrix}. \quad (6.4)$$

Now we can readily verify the claims about $\text{rank}(K)$. It remains to estimate the condition number $\kappa(K) = \|K\| \|K^{-1}\| = \|\bar{K}\| \|\bar{K}^{-1}\|$ provided that $r = n - \rho$ and the matrix K is nonsingular. To bound the norm $\|K\|$, note that $K = \begin{pmatrix} W & O_{r,n} \\ O_{n,r} & A \end{pmatrix} + \begin{pmatrix} O_{r,n} & I_{r,r} \\ I_{n,n} & O_{n,r} \end{pmatrix} \begin{pmatrix} -U & O_{n,n} \\ O_{r,r} & V^T \end{pmatrix}$, recall that $\|W\| \leq \|A\| = 1$, apply bound (2.7) and obtain

$$\|K\| \leq 1 + \max\{\|U\|, \|V\|\}.$$

By virtue of randomized bounds of Theorem 3.2 we expect to have the norms $\|U\|$ and $\|V\|$ in $O(1)$, that is bounded by a constant, and then $\|K\|$ is in $O(1)$ as well.

We conclude the proof by estimating the norm $\|\bar{K}\| = \|K\|$. We readily verify that

$$\bar{K}^{-1} = \begin{pmatrix} O_{r,r} & O_{r,n-r} & -U_1^{-1} \\ O_{n-r,r} & \Sigma_{n-r}^{-1} & -\Sigma_{n-r}^{-1} U_0 U_1^{-1} \\ V_1^{-T} & -V_1^{-T} V_0^T \Sigma_{n-r}^{-1} & V_1^{-T} (W + V_0^T \Sigma_{n-r}^{-1} U_0) U_1^{-1} \end{pmatrix}.$$

Apply bound (2.7) and deduce that

$$\|K^{-1}\| = \|\bar{K}^{-1}\| \leq N_1 + N_2 + N_3$$

where $N_1 = \max\{\|V_1^{-T}\|, \|\Sigma_{n-r}^{-1}\|, \|U_1^{-1}\|\}$, $N_2 = \max\{\|V_1^{-T} V_0^T \Sigma_{n-r}^{-1}\|, \|\Sigma_{n-r}^{-1} U_0 U_1^{-1}\|\}$ and $N_3 = \|V_1^{-T} (W + V_0^T \Sigma_{n-r}^{-1} U_0) U_1^{-1}\|$. Recall that $\|W\| \leq 1$, $\|\Sigma_{n-r}^{-1}\| = 1/\sigma_{n-r}(A)$, $\|V_0^T\| = \|V_0\|$, and $\|V_1^{-T}\| = \|V_1^{-1}\|$ and deduce that

$$N_1 = \max\{\|V_1^{-1}\|, \|U_1^{-1}\|, 1/\sigma_{n-r}(A)\},$$

$$N_2 \leq \max\{\|V_1^{-1}\| \|V_0\|, \|U_0\| \|U_1^{-1}\|\} / \sigma_{n-r}(A),$$

$$N_3 \leq \|V_1^{-1}\| \|U_1^{-1}\| (1 + \|V_0\| \|U_0\| / \sigma_{n-r}(A)).$$

Apply Theorems 3.1 and 3.2 to estimate the norms $\|U\|$, $\|V\|$, $\|V_0\|$, $\|U_0\|$, $\|U_1^{-1}\|$ and $\|V_1^{-1}\|$. Combine all the above bounds to estimate the norm $\|K^{-1}\|$. \square

Next we extend Theorem 6.4 to the case where $r \geq n - \rho$ and where we require that the leading block W be normalized, square and well conditioned.

Theorem 6.5. *Theorem 6.4 still holds where $r > n - \rho$ and the leading block W of the matrix K is normalized, square and well conditioned.*

Proof. Clearly parts (i) and (ii) of Theorem 6.4 are extended, and moreover we immediately deduce that for $r > n - \rho$ the matrix K is nonsingular with probability 1 and that its norm $\|K\|$ is expected to be in $O(1)$. It remains to estimate the norm $\|K^{-1}\|$. Assume SVDs $W = S_W \Sigma_W T_W^T$ and $A = S_A \Sigma_A T_A^T$, write $\bar{K} = \text{diag}(S_W^T, S_A^T) K \text{diag}(T_W, T_A)$ and observe that $\bar{K} = \begin{pmatrix} \Sigma_W & \bar{V}^T \\ \bar{U} & \Sigma_A \end{pmatrix}$ where $\bar{U}, \bar{V} \in \mathcal{G}_{0,1}^{\rho \times r}$ by virtue of Lemma 4.3, because the matrices S_W , T_W , S_A , and T_A are square and orthogonal. Now complete the proof by extending the techniques used in Section 5.2 in the proof of part (iv) of Theorem 1.1, in the case where $r > n - \rho$. \square

Corollary 6.2. *Keep the assumptions of Theorem 6.5, except that now let the matrix A have numerical rank ρ , rather than rank ρ . Then Theorem 6.5 is extended and furthermore the matrix K is expected to be well conditioned.*

Proof. We immediately extend parts (i) and (ii) of Theorem 6.4. To extend part (iii) truncate the SVD of the matrix A by setting to 0 all its singular values except for the ρ largest ones to obtain matrices $A - E \approx A$ of rank ρ and $\hat{K} = \begin{pmatrix} W & V^T \\ -U & A - E \end{pmatrix}$ such that $\|\hat{K} - K\| \leq \|E\| = \sigma_{\rho+1}(A)$.

The value $\sigma_{\rho+1}(A)$ is small because the matrix A has numerical rank ρ , whereas the norm $\|\hat{K}^{-1}\|$ is not expected to be large by virtue of Theorem 6.5. Therefore we can expect that $\|E\hat{K}^{-1}\| \leq 1/3$, and consequently that $\|K^{-1}\| \leq 1.5\|\hat{K}^{-1}\|$ by virtue of Theorem 2.3. Consequently the condition number $\kappa(K)$ is also expected to be of at most order $\|A\|/\sigma_{l-r}(A) = 1/\sigma_{n-r}(A)$ as in part (iii) of Theorem 6.5, but in the corollary this means that the matrix is well conditioned because we assume that the matrix A has numerical rank ρ , and so the ratio $\|A\|/\sigma_{\rho}(A)$ is not large. \square

The corollary implies that the matrix K is nonsingular with probability 1 and is expected to be well conditioned in the case where the matrix A has a numerical rank at least $n - r$. Let us extend our analysis to the case of rectangular matrices $A \in \mathbb{R}^{m \times n}$.

Theorem 6.6. *Suppose m , n , and r are three positive integers, $l = \min\{m, n\}$, $A \in \mathbb{R}^{m \times n}$, the matrix A has a numerical rank $\rho < l - r$, $U \in \mathcal{G}_{0,1}^{m \times r}$, $V \in \mathcal{G}_{0,1}^{n \times r}$, $W \in \mathbb{R}^{r \times r}$, $\|W\| \leq 1$, K is the $(m + r) \times (n + r)$ matrix defined by equation (6.1). Then (i) this matrix is rank deficient or ill conditioned if $\rho < l - r$, but otherwise has full rank with probability 1 and (ii) is expected to be well conditioned.*

Proof. Part (i) is readily verified. Let us prove part (ii). Suppose $A = S_A \Sigma_A T_A^T$ is the SVD of (2.5) and write $\bar{K}' = \text{diag}(I_r, S_A^T) K \text{diag}(I_r, T_A) = \begin{pmatrix} W & \bar{V}^T \\ \bar{U} & \Sigma_A \end{pmatrix}$ where $\bar{U} = -S_A^T U$ and $\bar{V}^T = V^T T_A$. Observe that $\kappa(K) = \kappa(\bar{K}')$, $\bar{U} \in \mathcal{G}_{0,1}^{m \times r}$, $\bar{V} \in \mathcal{G}_{0,1}^{n \times r}$, and $\kappa(K) = \kappa(\bar{K}')$ because S_A and T_A are square orthogonal matrices. Define the leading $l \times l$ submatrix $\hat{K}' = I_{m+r, n+r} \bar{K}' I_{n+r, m+r}^T$ for $I_{g,h}$ of (2.1) and observe that $\bar{K}' = \begin{pmatrix} \hat{K}' \\ \hat{U} \end{pmatrix}$, $\hat{U} = (U_2 \mid O_{m-n, n})$, $U_2 \in \mathcal{G}_{0,1}^{(m-n) \times q}$ if $m \geq l = n$, whereas $\bar{K}' = (\hat{K}' \mid \hat{V}^T)$, $\hat{V}^T = (V_2^T \mid O_{n-m, m})$, $V_2 \in \mathcal{G}_{0,1}^{(n-m) \times r}$ if $n \geq l = m$. Clearly $\sigma_l(\bar{K}') \geq \sigma_l(\hat{K}')$ (cf. Fact 2.3) and $\|\bar{K}'\| \leq \|\hat{K}'\| + \|F\|$ for $F = U_2$ or $F = V_2$. In both cases $F \in \mathcal{G}_{0,1}^{|n-m| \times r}$, and so we can expect that $\|\bar{K}'\| = O(\|\hat{K}'\|)$ because $\|\hat{K}'\| \geq \|\Sigma_A\| = \|A\| = 1$. Corollary 6.2 implies that with probability 1 the $(l + r) \times (l + r)$ matrix \hat{K}' is nonsingular, and then $\text{rank}(K) = \text{rank}(\bar{K}') = \text{rank}(\hat{K}') = l + r$, implying that the matrix K has full rank. Furthermore Corollary 6.2 implies that the matrix \hat{K}' is expected to be well conditioned. It remains to extend this property to the matrix K . Recall that $\kappa(\hat{K}') = \|\hat{K}'\|/\sigma_l(\hat{K}')$ and $\kappa(K) = \kappa(\bar{K}') = \|\bar{K}'\|/\sigma_l(\bar{K}')$ and combine the above equations with the bounds $\sigma_l(\bar{K}') \geq \sigma_l(\hat{K}')$ and $\|\bar{K}'\| = O(\|\hat{K}'\|)$, deduced earlier. \square

6.5 A randomized Toeplitz solver

Let us apply Theorem 2.6 to support randomized augmentation for solving a nonsingular Toeplitz linear system $T\mathbf{y} = \mathbf{b}$ of n equations provided the matrix T has numerical nullity 1.

To compute the vector $\mathbf{y} = T^{-1}\mathbf{b}$, we first embed the matrix T into a Toeplitz $(n + 1) \times (n + 1)$ matrix $K = \begin{pmatrix} w & \mathbf{v}^T \\ \mathbf{f} & T \end{pmatrix}$. We write $w = \mathbf{e}_1^T T \mathbf{e}_1$ and fill the vectors $\mathbf{f} = (f_i)_{i=1}^n$ and $\mathbf{v} = (v_i)_{i=1}^n$ with appropriate entries of the matrix T except for the two coordinates f_n and v_n , which we choose at random and then scale to have the ratios $\frac{|f_n|}{\|\mathbf{f}\|}$ and $\frac{|v_n|}{\|\mathbf{v}\|}$ neither large nor small.

Part (b) of Theorem 2.6 expresses the inverse T^{-1} via the vectors $\mathbf{v} = K^{-1}\mathbf{e}_1$ and $\mathbf{w} = K^{-1}\mathbf{e}_{n+1}$.

In view of Section 3.2 and Appendix A, this policy is likely to produce a nonsingular matrix K whose inverse is likely to have a nonzero entry $\mathbf{e}_1^T K^{-1} \mathbf{e}_1$. In good accordance with these formal results our tests have always produced nonsingular and well conditioned matrices K such that $\mathbf{e}_1^T K^{-1} \mathbf{e}_1 \neq 0$.

To summarize, we reduce the solution of a nonsingular ill conditioned Toeplitz linear system $T\mathbf{y} = \mathbf{b}$ to computing highly accurate solutions of two linear systems $K\mathbf{x} = \mathbf{e}_1$ and $K\mathbf{z} = \mathbf{e}_{n+1}$, both expected to be well conditioned. High accuracy shall counter the magnification of the input and rounding errors, expected in the case of ill conditioned input.

In the important special case where a Toeplitz matrix T is real symmetric, we choose real scalars w and $f_n = v_n$ to yield a real symmetric matrix $K = \begin{pmatrix} w & \mathbf{v}^T \\ \mathbf{v} & T \end{pmatrix}$. In this case $J_{n+1}K^{-1}J_{n+1} = K^{-1}$, and so $K^{-1}\mathbf{e}_{n+1} = J_{n+1}K^{-1}\mathbf{e}_1$ because $J_{n+1}\mathbf{e}_{n+1} = \mathbf{e}_1$. Thus we only need to solve a single linear system with the matrix K . For the transition back to the solution of the original problem, we can employ expression (6.2) or Theorem 2.6. Hereafter we refer to the resulting algorithm for the linear system $T\mathbf{y} = \mathbf{b}$ as **Algorithm 6.1**. In Section 10.6 we test this algorithm for solving an ill conditioned real symmetric Toeplitz linear system.

One can readily extend the approach of this section to the case of Toeplitz-like, Hankel and Hankel-like inputs and to augmenting the input matrix with r rows and r columns for $r > 1$.

7 Low-rank approximation, approximation of singular spaces, and computation of numerical rank

7.1 Randomized low-rank approximation: an outline and an extension to approximation by structured matrices

Our next theorem expresses a rank- ρ approximation to a matrix A through an approximate matrix basis for the left or right leading singular space $\mathbb{T}_{\rho,A}$ or $\mathbb{S}_{\rho,A}$. We can obtain such basis by computing the SVD of the matrix A or its rank-revealing factorization [GE96], [HP92], [P00a], but if the matrix A has a small numerical rank ρ and if we are given its reasonably small upper bound ρ_+ , then with a probability near 1 we can compute such basis at a low cost from the product $A^T G$ for $G \in \mathcal{G}_{0,1}^{m \times \rho_+}$. Theorem 7.2 of Section 7.3 formally supports correctness of the respective randomized algorithm, but our tests support it consistently even where $G \in \mathcal{T}_{0,1}^{m \times \rho_+}$ (see Tables 10.9 and 10.10), and we conjecture that the same is true for various other classes of sparse and structured matrices G defined by fewer random parameters. We specify a low-rank approximation algorithm in Section 7.4, its amendments in Section 7.7, and some related randomized algorithms of independent interest for the approximation of leading and trailing singular spaces of an ill conditioned matrix in Sections 7.5 and 7.6. By applying low-rank approximation algorithms to a displacement of a matrix W having a possibly unknown numerical displacement rank d , that is lying near some matrices with displacement rank d , we can approximate the matrix W by one of these matrices and output d as by-product. In Section 8.3 we apply this observation to Newton's structured matrix inversion.

7.2 The first basic theorem: low-rank approximation via the basis of a leading singular space

The following theorem expresses a rank- q approximation (within an error norm $\sigma_{q+1}(A)$) to a matrix A through a matrix basis of its leading singular space $\mathbb{T}_{q,A}$ or $\mathbb{S}_{q,A}$.

Theorem 7.1. *Suppose A is an $m \times n$ matrix, $S_A \Sigma_A T_A^T$ is its SVD of (2.5), q is a positive integer, $q \leq \min\{m, n\}$, and T and S are matrix bases for the spaces $\mathbb{T}_{q,A}$ and $\mathbb{S}_{q,A}$, respectively. Then*

$$\|A - AT(T^T T)^{-1}T^T\| = \|A - S(S^T S)^{-1}S^T A\| = \sigma_{q+1}(A). \quad (7.1)$$

For orthogonal matrices T and S we have $T^T T = S^T S = I_q$ and

$$\|A - ATT^T\| = \|A - SS^T A\| = \sigma_{q+1}(A). \quad (7.2)$$

Proof. Let us first write $P = T_{q,A}T_{q,A}^T$ and $r = n - q$ and estimate the norm $\|A - AP\|$. We have $AP = S_A \Sigma_A T_A^T T_{q,A} T_{q,A}^T$. Substitute $T_A^T T_{q,A} = \begin{pmatrix} I_q \\ O_{r,q} \end{pmatrix}$ and obtain $AP = S_A \Sigma_A \begin{pmatrix} T_{q,A}^T \\ O_{r,q} \end{pmatrix}$, whereas $A = S_A \Sigma_A \begin{pmatrix} T_{q,A}^T \\ T_{A,r}^T \end{pmatrix}$. Therefore

$$A - AP = S_A \Sigma_A \begin{pmatrix} O_{q,n} \\ T_{A,r}^T \end{pmatrix} = S_A \text{diag}(O_q, \text{diag}(\sigma_j)_{j=q+1}^n) \begin{pmatrix} O_{q,n} \\ T_{A,r}^T \end{pmatrix},$$

and so $\|A - AP\| = \|\text{diag}(\sigma_j)_{j=q+1}^n\| = \sigma_{q+1}$ because S_A and $T_{A,r}$ are orthogonal matrices. Similarly deduce that $\|A - S_{q,A} S_{q,A}^T A\| = \sigma_{q+1}(A)$. This proves (7.1) and (7.2) for $T = T_{q,A}$ and $S = S_{q,A}$.

Now let the matrices T and S have full rank, $\mathcal{R}(T) = \mathbb{T}_{q,A} = \mathcal{R}(T_{q,A})$, $\mathcal{R}(S) = \mathbb{S}_{q,A} = \mathcal{R}(S_{q,A})$, and so $T = T_{q,A}U$ and $S = S_{q,A}V$ for two nonsingular matrices U and V . Consequently $T(T^T T)^{-1} T^T = T_{q,A}U(U^T T_{q,A}^T T_{q,A}U)^{-1} U^T T_{q,A}^T$. Substitute $T_{q,A}^T T_{q,A} = I_q$ and deduce that $(U^T T_{q,A}^T T_{q,A}U)^{-1} = (U^T U)^{-1} = U^{-1}U^{-T}$. Therefore $U(U^T T_{q,A}^T T_{q,A}U)^{-1} U^T = U U^{-1} U^{-T} U^T = I_q$, and so $T(T^T T)^{-1} T^T = T_{q,A}U(U^T T_{q,A}^T T_{q,A}U)^{-1} U^T T_{q,A}^T = T_{q,A} T_{q,A}^T$. Similarly $S(S^T S)^{-1} S^T = S_{q,A} S_{q,A}^T$, implying the desired extension. \square

7.3 The second basic theorem: a basis of a leading singular space via randomized products

The following theorem supports randomized approximation of matrix bases for the leading singular spaces $\mathbb{T}_{\rho,A}$ and $\mathbb{S}_{\rho,A}$ of a matrix A having numerical rank ρ .

Theorem 7.2. *Suppose a matrix $A \in \mathbb{R}^{m \times n}$ has a numerical rank ρ , $H \in \mathcal{G}_{0,1}^{n \times \rho_+}$ and $G \in \mathcal{G}_{0,1}^{m \times \rho_+}$ for $\rho_+ \geq \rho$. Then the matrices $T = A^T G$ and $S = AH$ have full rank with probability 1 and we can expect that they have numerical rank ρ and that*

$$S + \Delta = S_{\rho,A}U \text{ and } T + \Delta' = T_{\rho,A}V \quad (7.3)$$

for two matrices Δ and Δ' having norms of order $\sigma_{\rho+1}(A)$ and for two nonsingular matrices U and V having condition numbers of at most order $\|A\|/(\sigma_\rho(A)\sqrt{\rho})$.

Proof. The techniques of Section 3.2 and Theorem 4.1 support the claims about ranks and numerical ranks. It remains to deduce the former probabilistic relationship $\mathbb{S}_{\rho,AH+\Delta} = \mathbb{S}_{\rho,A}$ of (7.3) because we can apply it to A^T to obtain the latter relationship $\mathbb{T}_{\rho,A^T G+\Delta'} = \mathbb{T}_{\rho,A}$.

Assume the SVD $A = S_A \Sigma_A T_A^T$ and note that $\|\Sigma_A - \text{diag}(\Sigma_{\rho,A}, O_{m-\rho, n-\rho})\| \leq \sigma_{\rho+1}(A)$. Consequently $\|A - S_A \text{diag}(\Sigma_{\rho,A}, O_{m-\rho, n-\rho}) T_A^T\| \leq \sigma_{\rho+1}(A)$ and $AH = S - \Delta$, $S = S_{\rho,A}U$, $\|\Delta\| \leq \sigma_{\rho+1}(A) \|AH\|$ where $U = \Sigma_{\rho,A}B$, $B = T_{\rho,A}^T H$, and we can expect that the norm $\|H\|$ is bounded from above and below by two positive constants (see Theorem 3.2). This implies (7.3). It remains to estimate $\kappa(U)$.

With probability 1 the $\rho \times \rho$ matrices B and U are nonsingular (see Section 3.2). Furthermore we have $\|U\| \leq \|\Sigma_{\rho,A}\| \|B\|$ where $\|\Sigma_{\rho,A}\| = \|A\|$ and $\|B\| \leq \|T_{\rho,A}\| \|H\| = \|H\|$. So $\|U\| \leq \|A\| \|H\| = O(\|A\|)$. We also have $\|U^+\| \leq \|\Sigma_{\rho,A}^{-1}\| \|B^{-1}\|$ for nonsingular matrix B . Observe that $\|\Sigma_{\rho,A}^{-1}\| = 1/\sigma_\rho(A)$, apply Theorem 4.1 where $M = T_{\rho,A}^T$, $\hat{r} = \rho$ and $\sigma_{r(M)}(M) = \sigma = 1$ and obtain that the norm $\|B^{-1}\|$ is expected to have at most order $1/\sqrt{\rho}$. Summarizing we can expect that the norm $\|U^+\|$ has at most order $1/(\sigma_\rho(A)\sqrt{\rho})$. Consequently $\kappa(U) = \|U\| \|U^+\|$ has at most order $\|A\|/(\sigma_\rho(A)\sqrt{\rho})$. \square

7.4 A prototype algorithm for low-rank approximation

Theorems 7.1 and 7.2 imply correctness of the following prototype algorithm (cf. [HMT11, Section 10.3]), where the input matrix has an unknown numerical rank and we know its upper bound.

Proto-Algorithm 7.1. Rank- ρ approximation of a matrix.

INPUT: A matrix $A \in \mathbb{R}^{m \times n}$ having an unknown numerical rank ρ , an integer $\rho_+ \geq \rho$, and two tolerances τ and τ' of order $\sigma_{\rho+1}(A)/\|A\|$. (We can choose τ at Stage 2 based on rank revealing factorization, can choose τ' at Stage 3 based on the required output accuracy, and can adjust both tolerances if the algorithm fails to produce a satisfactory output.)

OUTPUT: FAILURE (with a low probability) or an integer ρ and two matrices $T \in \mathbb{R}^{n \times \rho}$ and $A_\rho \in \mathbb{R}^{m \times n}$, both having ranks at most ρ and such that $\|A_\rho - A\| \leq \tau' \|A\|$ and T satisfies (7.3) for $\|\Delta'\| \leq \tau \|A\|$.

COMPUTATIONS:

1. Compute the $n \times \rho_+$ matrix $T' = A^T G$ for $G \in \mathcal{G}_{0,1}^{m \times \rho_+}$.
2. Compute a rank revealing factorization of the matrix T' and choose the minimal integer s and an $n \times s$ matrix T such that $\|T' - (T \mid O_{n, \rho_+ - s})\| \leq \tau \|A\|$.
3. Compute the matrix $A_s = AT(T^T T)^{-1} T^T$. Output $\rho = s$, T and A_ρ and stop if $\|A_\rho - A\| \leq \tau' \|(A)\|$. Otherwise output FAILURE and stop.

Assume that both tolerances τ and τ' have been chosen properly. Then by virtue of Theorem 7.2, we can expect that at Stage 2 $s = \rho$ and T is an approximate matrix basis for the singular space $\mathbb{T}_{\rho, A}$ (within an error norm of at most order $\sigma_{\rho+1}(A)$). Consequently Stage 3 outputs FAILURE with a probability near 0, by virtue of Theorems 7.1. (In the case of FAILURE we can reapply the algorithm for new values of random parameters or for the adjusted tolerance values τ and τ' .) At Stage 2 we have $s \leq \rho$ because $\text{nrnk}(A^T G) \leq \text{nrnk}(A) = \rho$, whereas the bound $\|A_\rho - A\| \leq \tau' \|(A)\|$ at Stage 3 implies that $s \geq \text{nrnk}(A)$. This certifies the outputs ρ , T , and A_ρ of the algorithm.

We can similarly approximate the matrix A by a rank- ρ matrix $S(S^T S)^{-1} S^T A$, by first computing the matrix $S' = AH$ for $H \in \mathcal{G}_{0,1}^{n \times \rho_+}$, then computing its rank revealing factorization, which is expected to define an approximate matrix basis S for the space $\mathbb{S}_{\rho, A}$, and finally applying Theorem 7.1. We have $T^T T = I_n$ and $S^T S = I_m$ where the matrices T and S are orthogonal, and then the expressions for rank- ρ approximation are simplified.

Remark 7.1. One can weaken reliability of the output to simplify Stage 3 by testing whether $\|K^T(A - A_\rho)L\| \leq \tau \|K\| \|A\| \|L\|$ for matrices $K \in \mathcal{G}_{0,1}^{m \times \rho'}$ and $L \in \mathcal{G}_{0,1}^{n \times \rho''}$ and for two small positive integers ρ' and ρ'' , possibly for $\rho' = \rho'' = 1$, instead of testing whether $\|A_\rho - A\| \leq \tau' \|(A)\|$. One can similarly simplify Stage 2.

Remark 7.2. For $\rho_+ = \rho$ Stage 2 can be omitted because the matrix $A^T G$ is expected to be a desired approximate matrix basis by virtue of Theorem 7.2. The increase of the dimension ρ_+ beyond ρ (called oversampling in [HMT11]) is relatively inexpensive if the bound ρ_+ is small. [HMT11] suggests using small oversampling even if the numerical rank ρ is known, because we have

$$\text{Probability } \{\|A - ATT^T\| \leq (1 + 9\sqrt{\rho_+ \min\{m, n\}})\sigma_{\rho+1}(A)\} \geq 1 - 3(\rho_+ - \rho)^{\rho - \rho_+} \text{ for } \rho_+ > \rho.$$

Theorem 7.2, however, bounds the norm $\|A - ATT^T\|$ strongly also for $\rho = \rho_+$, in good accordance with the data of Tables 10.9 and 10.10.

7.5 Computation of nmbs and approximation of trailing singular spaces

One can approximate trailing singular spaces $\mathbb{S}_{A, \rho}$ and $\mathbb{T}_{A, \rho}$ of a nonsingular ill conditioned matrix A having numerical rank ρ by applying Proto-Algorithm 7.1 to the matrix A^{-1} , because $\mathbb{T}_{\rho, A^{-1}} = \mathbb{S}_{A, \rho}$ and $\mathbb{S}_{\rho, A^{-1}} = \mathbb{T}_{A, \rho}$. Next we achieve the same goal without inverting the matrix A , furthermore we cover the case of rectangular inputs. At first we compute a nmb of a rank deficient matrix A and then approximate the trailing singular space $\mathbb{T}_{A, r}$ of an ill conditioned matrix A by truncating its SVD and applying Theorem 2.2 or 2.3. We can compute left nmbs and approximate left trailing singular spaces by applying the same algorithms to the matrix A^T .

Theorem 7.3. [PQ10, Theorem 3.1 and Corollary 3.1]. Suppose a matrix $A \in \mathbb{R}^{m \times n}$ has rank ρ , $U \in \mathbb{R}^{m \times r}$, $V \in \mathbb{R}^{n \times r}$, and the matrix $C = A + UV^T$ has full rank n . Write $B = C^{(I)}U$. Then $r \geq n - \rho$, $\mathcal{R}(B) \supseteq \mathcal{N}(A)$; moreover if $r = n - \rho$, then $C^{(I)}U = \text{nmb}(A)$. Furthermore $\mathcal{R}(BX) = \mathcal{N}(A)$ if $\mathcal{R}(X) = \mathcal{N}(AB)$. (Note that $AB = U(I_r V^T C^{-1} U)$ for $m = n$.)

Theorem 7.4. Suppose that $A \in \mathbb{R}^{m \times n}$, $U \in \mathbb{R}^{m \times q}$, $V \in \mathbb{R}^{n \times s}$, $W \in \mathbb{R}^{s \times q}$, $K = \begin{pmatrix} W & V^T \\ -U & A \end{pmatrix}$,

$\text{rank}(W) = q \geq \text{nul}(A)$, $\text{rank}(K) = n + q$, $m \geq n$. Write $Y = (O_{n, q} \mid I_n)K^{(I)} \begin{pmatrix} O_{s, q} \\ U \end{pmatrix}$. Then

(a) $\mathcal{N}(A) \subseteq \mathcal{R}(Y)$ and if $\text{rank}(U) = \text{nul}(A)$, then $\mathcal{N}(A) = \mathcal{R}(Y)$,

- (b) $\mathcal{R}(YZ) = \mathcal{N}(A)$ if $\mathcal{R}(Z) = \mathcal{N}(AY)$, whereas
(c) $\mathcal{R}(Z) = \mathcal{N}(AY)$ if $\mathcal{R}(YZ) = \mathcal{N}(A)$ and if $\text{rank}(Y) = q$.

Proof. See [PQa, Theorems 11.2 and 11.3]. \square

Remark 7.3. Both theorems define aggregation processes (cf. [MP80]). For $r > n - \rho$, Theorem 7.3 reduces the computation of a $\text{nmb}(A)$ to the same task for the input BX of a smaller size $n \times (r - n + \rho)$. Furthermore, suppose that the matrices U and Y have full rank q . Then part (a) of Theorem 7.4 implies that Y is a $\text{nmb}(A)$ if $q = \text{nul}(A)$, but otherwise parts (b) and (c) reduce the original task of computing a $\text{nmb}(A)$ to the case of the input AY of a smaller size $m \times (q - \text{nul}(A))$.

Theorem 7.5. Assume that $U \in \mathbb{R}^{m \times r_+}$, $V \in \mathbb{R}^{n \times r_+}$, $m \geq n$, a real $m \times n$ matrix A has numerical rank $\rho = n - r$, and the matrix $C = A + UV^T$ has full rank and is well conditioned. Then $\rho \geq n - r_+$ and there is a scalar c independent of A, U, V, m, n and ρ such that $\|C^+UX - T_{A,r}\| \leq c\sigma_{\rho+1}(A)\|U\|$ where $X \in \mathbb{R}^{r_+ \times r}$, $X = \text{nmb}(AC^+U + \Delta)$, $\|\Delta\| \leq c\sigma_{\rho+1}(A)\|U\|$.

Proof. The theorem turns into Theorem 7.3 if $\rho = \text{nrank}(A) = \text{rank}(A)$. If $\rho = \text{nrank}(A) < \text{rank}(A)$, set to zero all but the ρ largest singular values in the SVD of the matrix A . Then $\rho = \text{nrank}(A - E) = \text{rank}(A - E)$ and the theorem holds for the resulting matrix $A - E$ and the matrix $C - E = A - E + UV^T$. Therefore $T_{A-E,r} = (C - E)^+UX$ where for X we choose an orthogonal $\text{nmb}((A - E)((C - E)^+U))$, of size $r_+ \times r$. Clearly $\|T_{A-E,r} - T_{A,r}Q\| = O(\sigma_{\rho+1}(A))$ for some $r \times r$ orthogonal matrix Q , and it remains to estimate the norm $\|(C - E)^+UX - C^+UX\|$. We have $\|((C - E)^+ - C^+)UX\| \leq \|(C - E)^+ - C^+\| \|U\|$. The norm $\|E\| = \sigma_{\rho+1}(A)$ is small because the matrix A has numerical rank ρ , whereas the norm $\|(C - E)^+\|$ is not large because the full rank matrix C is well conditioned. Therefore the value $\tau = \|(C - E)^+\| - \|C^+\|$ has at most order $\sigma_{\rho+1}(A)$ by virtue of Theorem 2.2. \square

Corollary 7.1. Suppose a normalized real $m \times n$ matrix A has numerical rank $\rho = n - r$, $U \in \mathcal{G}_{0,1}^{m \times r_+}$, $V \in \mathcal{G}_{0,1}^{n \times r_+}$, $m \geq n$, and $C = A + UV^T$. Then (i) the matrix C is singular or ill conditioned if $r_+ < r$ but otherwise (ii) has full rank with probability 1, and (iii) we can expect that the matrix C^+UX is an approximate matrix basis for the singular space $\mathbb{T}_{A,r}$ within an error norm of at most order $\sigma_{\rho+1}(A)$ where X is an orthogonal $\text{nmb}(AC^+U + \Delta)$ of the size $r_+ \times r$ and $\|\Delta\| \leq c\sigma_{\rho+1}(A)$.

Proof. Part (i) is immediately verified. Furthermore by virtue of Theorem 5.6 the matrix C has full rank with probability 1 and is expected to be well conditioned, whereas the norm $\|U\|$ is expected to be not large by virtue of Theorem 3.2. Therefore Corollary 7.1 follows from Theorem 7.5. \square

Likewise by employing Theorems 6.3 and 7.4 instead of Theorems 5.6 and 7.3, we obtain the following result.

Corollary 7.2. Suppose that a normalized real $m \times n$ matrix A has numerical nullity $r = \text{nnul}(A)$, $U \in \mathcal{G}_{0,1}^{m \times q}$, $V \in \mathcal{G}_{0,1}^{n \times s}$, $W \in \mathcal{G}_{0,1}^{s \times q}$, $K = \begin{pmatrix} W & V^T \\ -U & A \end{pmatrix}$, $\text{rank}(W) = q$, $\text{rank}(K) = n + q$, $Y = (O_{n,q} \mid I_n)K^+ \begin{pmatrix} O_{s,q} \\ U \end{pmatrix}$, and $m \geq n$. Then (i) the matrix K is rank deficient or ill conditioned where $q < r$ but otherwise has full rank with probability 1 and is expected to be well conditioned. Furthermore we can expect that within an error norm of at most order $\sigma_{n-q+1}(A)$ a matrix basis for the singular space $\mathbb{T}_{A,q}$ is approximated by (ii) the matrix Y if $r = q$ or (iii) the matrix YZ if $q > r$ where $Z \in \mathbb{R}^{q \times r}$, $Z = \text{nmb}(AY + \Delta)$, $\|\Delta\| \leq c\sigma_{n-q+1}(A)$.

Corollaries 7.1 and 7.2 (for $s = q$) imply correctness of the two following Prototype Algorithms.

Proto-Algorithm 7.2. An approximate basis for a trailing singular space by using randomized additive preprocessing.

INPUT: A matrix $A \in \mathbb{R}^{m \times n}$ for $m \geq n$ with $\|A\| \approx 1$, an upper bound r_+ on its unknown numerical nullity $r = \text{nnul}(A)$, and two tolerances τ and τ' of order $\sigma_{n-r+1}(A)$. (The tolerances are defined by the requested output accuracy. In a variation of the algorithm one can reapply it with a decreased tolerance τ' instead of outputting FAILURE at Stage 4.)

OUTPUT: *FAILURE* (with a low probability) or the numerical nullity r and an approximate matrix basis B , within an error norm in $O(\sigma_{n-r+1}(A))$, of the trailing singular space $\mathbb{T}_{A,r}$.

INITIALIZATION: Generate two matrices $U \in \mathcal{G}_{0,1}^{m \times r_+}$ and $V \in \mathcal{G}_{0,1}^{n \times r_+}$ for σ of order $\|A\|$.

COMPUTATIONS:

1. Compute the matrix $C = A + UV^T$.
2. Stop and output *FAILURE* if this matrix is rank deficient or ill conditioned. Otherwise compute the matrices $Y = C^+U$ and AY .
3. Output $r = r_+$ and $B = Y$ and stop if $\|AY\| \leq \tau\|A\| \|Y\|$.
4. Otherwise apply an algorithm (e.g. employing SVD, rank revealing factorization, a technique from [PQ10] or [PQa], or one of Proto-Algorithms 7.2 and 7.3) that for the matrix AY and a fixed tolerance τ' computes an integer r and an orthogonal approximate matrix basis X (of size $r_+ \times r$) for the space $\mathbb{T}_{AY,r}$. If $\|AB\| \leq \tau\|A\| \|B\|$ then output r and $B = YX$ and stop. Otherwise output *FAILURE* and stop.

Proto-Algorithm 7.3. An approximate basis for a trailing singular space by using randomized augmentation.

INPUT, OUTPUT and Stages 3 and 4 of COMPUTATIONS are as in Proto-Algorithm 7.2.

INITIALIZATION: Generate three matrices $U \in \mathcal{G}_{0,1}^{m \times r_+}$, $V \in \mathcal{G}_{0,1}^{n \times r_+}$, and $W \in \mathcal{G}_{0,1}^{r_+ \times r_+}$ for σ of order $\|A\|$.

COMPUTATIONS:

1. Stop and output *FAILURE* if the matrix $K = \begin{pmatrix} W & V^T \\ -U & A \end{pmatrix}$ is rank deficient or ill conditioned.
2. Otherwise compute the matrices $Y = (O_{n,r_+} \mid I_n)K^+ \begin{pmatrix} O_{r_+, r_+} \\ U \end{pmatrix}$ and AY .

7.6 Alternative methods for the approximation of leading singular spaces

Next we extend Theorem 7.5 and Corollary 7.1 assuming a nonsingular input matrix and an upper bound on the numerical rank of its inverse.

Theorem 7.6. (Cf. Remark 7.4.) Assume that five matrices $A \in \mathbb{R}^{n \times n}$, $U_- \in \mathbb{R}^{n \times q_+}$, $H = I_{q_+} + V_-AU_-^T$, and $C_- = A - AU_-H^{-1}V_-^TA$ have full ranks, the matrix C_- is well conditioned, and $q = \text{nrnk}(A) \leq q_+$. Then there exists a scalar c_- independent of A , U_- , V_- , n and q_+ and such that $\|C_-^TV_-Y_- - T_{q,A}\| \leq c_- \sigma_{q+1}(A)$ where $Y_- \in \mathbb{R}^{q_+ \times q}$, Y_- is a matrix basis for the space $\mathbb{T}_{A^{-1}C_-^TV_+ \Delta, q}$ and $\|\Delta\| \leq c_- \sigma_{q+1}(A)$.

Proof. Recall that $C_-^{-1} = A^{-1} + U_-V_-^T$ (cf. (2.10)) and that $\text{nrnk}(A) = \text{nnul}(A^{-1})$. Rewrite SVDs $A = S_A \Sigma_A T_A^T$ and $A^{-1} = T_A \Sigma_A^{-1} S_A^T$ as follows,

$$A = (S_{q,A} \mid S_{A,n-q}) \text{diag}(\Sigma_{q,A}, \Sigma_{A,n-q})(T_{q,A} \mid T_{A,n-q})^T,$$

$$A^{-1} = (T_{q,A} \mid T_{A,n-q}) \text{diag}(\Sigma_{q,A}^{-1}, \Sigma_{A,n-q}^+)(S_{q,A} \mid S_{A,n-q})^T.$$

Apply Theorem 7.5 to the matrices A^{-1} , C^{-1} , and X replacing A , C , and Y , respectively. \square

Remark 7.4. One can first compute the numerical nullity $q = \text{nnul}(A)$ of an ill conditioned matrix A (see Section 8.2 on this computation) and then an approximate matrix basis C_-^TV of the space $\mathbb{T}_{q,A}$. This can be more attractive than computing the matrix $A^{-1}C_-^TV$. In the next two corollaries we assume that the numerical rank of the input matrix A is available.

Corollary 7.3. (Cf. Remark 7.4.) Suppose $A \in \mathbb{R}^{n \times n}$, $U_-, V_- \in \mathcal{G}_{0,\sigma}^{n \times q}$, σ has order $\|A^{-1}\|$, $H = I_q + V_- AU_-^T$, $C_- = A - AU_- H^{-1} V_-^T A$ for $H = I_q + V_-^T AU_-$ (cf. (2.10)), and $q = \text{nrnk}(A)$. (See Section 5.5 on estimating the norm $\|A^{-1}\|$.) Then the matrix $C_-^T V_-$ is expected to approximate within an error norm of at most order $\sigma_{q+1}(A^{-1})$ a matrix basis of the leading singular space $\mathbb{T}_{q,A}$.

Proof. By virtue of Theorem 5.6 the matrix C_- of Theorem 7.6 has full rank with probability 1 and is expected to be well conditioned, and so Corollary 7.3 follows from Theorem 7.6 for $q = q_+$. \square

The *dual augmentation* of the following corollary provides an alternative expression for an approximate matrix basis of a leading singular space.

Corollary 7.4. Suppose that $A \in \mathbb{R}^{n \times n}$, $U, V \in \mathcal{G}_{0,\sigma}^{n \times q}$, $W \in \mathcal{G}_{0,\sigma}^{q \times q}$, the matrix A is nonsingular, σ has order $\|A^{-1}\|$, $q = \text{nrnk}(A)$, $K_+ = \begin{pmatrix} W & V^T \\ -U & A^{-T} \end{pmatrix}$, $\text{rank}(W) = r$, and $\text{rank}(K_+) = n + q$. Then we can expect that the matrix $T_+ = (O_{n,q} \mid I_n) K_+^{-1} \begin{pmatrix} O_{q,q} \\ U \end{pmatrix}$ approximates within an error norm in $O(\sigma_{q+1}(A))$ a matrix basis for the right leading singular space $\mathbb{T}_{q,A}$.

Proof. Write SVD $A = S_A \Sigma_A T_A^T$ of (2.5) and deduce that $A^{-T} = S_A (\Sigma_A)^{-T} T_A^T$ where $(\Sigma_A)^{-T} = \text{diag}(1/\sigma_j(A))_{j=1}^n$. Note that $\text{nrnk}(A) = \text{nnul}(A^{-T})$ and apply Corollary 7.2 replacing the matrix A with A^{-T} . \square

Closer examination of the expression for the matrix T_+ enables us to simplify it as follows,

$$T_+ = B - BS^{-1}V^TB \text{ for } B = A^TU \quad (7.4)$$

where $S = W + U^T A^T V$. S^{-1} is the only matrix inverse involved into computing T_+ (cf. (2.10)).

7.7 Some amendments

Remark 7.5. Approximation of the leading and trailing singular spaces as well as the computation of numerical rank and numerical nullity (see Section 8.2) are facilitated as the gaps increase between the singular values of the input matrix A . This motivates using the power transforms $A \implies B_h = (AA^T)^h A$ for positive integers h because $\sigma_j(B_h) = (\sigma_j(A))^{2h+1}$ for all j .

Remark 7.6. In the case where $m = n$ the computations are simplified and stabilized, and furthermore we can apply Theorem 7.3 or 7.4 to both A and A^T to define both left and right nmbs. We can reduce to this case the computation for a rectangular matrix A in various ways, e.g., by observing that (a) $\mathcal{N}(A) = \mathcal{N}(A^T A)$, (b) $\mathcal{N}(A) = \mathcal{N}(B^T A)$ if $A, B \in \mathbb{R}^{m \times n}$ and the matrix B has full rank $m \leq n$, and (c) $(A \mid O_{n,m-n})\mathbf{u} = \mathbf{0}_m$ if and only if $A\hat{\mathbf{u}} = \mathbf{0}_m$ provided $m \geq n$ and $\hat{\mathbf{u}} = (I_n \mid O_{n,m-n})\mathbf{u}$, whereas $(A^T \mid O_{n,m-n})\mathbf{v} = \mathbf{0}_n$ if and only if $\hat{\mathbf{v}} = \mathbf{0}_n^T$ provided $m < n$ and $\hat{\mathbf{v}} = (I_m \mid O_{n-m,m})\mathbf{v}$. Furthermore given an $m \times n$ matrix A for $m > n$, we can represent it as the block vector $A = (B_1^T \mid B_2^T \mid \dots \mid B_h^T)^T$ where B_i are $k_i \times n$ blocks for $i = 1, \dots, h$, $\sum_{i=1}^h k_i = m$, and observe that $\mathcal{N}(A) = \cap_{i=1}^h \mathcal{N}(B_i)$, and we can compute the intersection of null spaces by applying [GL96, Theorem 12.4.1]. One can extend these comments to the tasks of the approximation of the singular spaces of ill conditioned matrices.

8 Sparse and structured randomization. Numerical rank without pivoting and orthogonalization

8.1 Randomized structured preprocessing

Would the additive preprocessing $A \implies C = A + UV^T$ preserve the structure of an $n \times n$ matrix A where $A \in \mathbb{R}^{m \times n}$, $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{n \times r}$? Adding the matrix UV^T makes small impact on the structure if the ratio $r/\min\{m, n\}$ is small, e.g., the displacement rank increases by $O(r)$ (cf.

[P01]), but we can control this impact even for large values r by endowing the matrices U and V with proper structure. Given a pair of standard Gaussian random Toeplitz $n \times r$ matrices U and V and a displacement generator of a small length d for a nonsingular ill conditioned Toeplitz-like $n \times n$ matrix A that has a numerical nullity $r = \text{nnul}(A)$ and a norm $\|A\| \approx 1$, we can readily compute a displacement generator of length $d + O(1)$ for the matrix $C = A + UV^T$. By exploiting the structure we can operate with this matrix in nearly linear arithmetic time, e.g., solve a nonsingular linear system $A\mathbf{y} = \mathbf{b}$ in $O(d^2 n \log^2 n)$ flops, even where r is large (see Theorem 2.9). Both randomized augmentation $A \Rightarrow K = \begin{pmatrix} W & V^T \\ -U & A \end{pmatrix}$ for proper choice of the random blocks U , V , and W and random sampling $A \Rightarrow A^T G$ and $A \Rightarrow AH$ for proper choice of random matrices G and H preserve matrix sparseness and structure even better. Empirically these maps preserve their preconditioning properties for such choices of the matrices G , H , U , V , and W ; likewise endowing random multipliers with sparseness and structure keeps the support for safe numerical GENP and block Gaussian elimination (see Remark 4.2 and Tables 10.5, 10.6, 10.10, and 10.16).

Remark 8.1. *Alternative deterministic techniques of homotopy continuation also support inversion in nearly linear time of nonsingular Toeplitz matrices and other matrices with displacement structure (see [P01, Section 6.9], [P07], [P10]).*

8.2 Numerical rank without pivoting and orthogonalization

If we know the numerical rank ρ of a matrix A , then we can simplify Proto-Algorithm 7.1 for rank- ρ approximation as well as the computation of approximate bases for the leading and trailing singular spaces of the matrix A (see Remarks 7.2 and 7.4 and Corollaries 7.1 and 7.2).

The customary algorithms for the numerical rank of a matrix rely on computing its SVD or rank revealing factorization, which involve pivoting and orthogonalization and thus destroy matrix sparseness and structure. Randomized Proto-Algorithm 7.1 is a noncostly alternative where the given upper bound ρ_+ on the numerical rank is small. Indeed Proto-Algorithm 7.1 uses rank revealing factorization at Stage 2 and matrix inversion or orthogonalization at Stage 3, but in these cases only deals with matrices of small sizes if ρ_+ is small.

Next we describe other alternatives that avoid pivoting and orthogonalization even where the numerical rank ρ is large. As by-product they compute an approximate matrix basis within an error norm in $O(\sigma_{\rho+1}(A))$ for the leading singular space $\mathbb{T}_{\rho,A}$ of an $m \times n$ matrix A and if we wish also rank- ρ approximation of the matrix A (see Remark 8.2). We let $m \geq n$ (else shift to A^T), let $[\rho_-, \rho_+] = [0, n]$ unless we know a more narrow range, and successively test the selected candidate integers in the range $[\rho_-, \rho_+]$ until we find the numerical rank ρ . To improve reliability, we can repeat the tests for distinct values of random parameters.

Exhaustive search defines and verifies the numerical rank ρ with probability near 1, but with proper policies one can use fewer and simpler tests because for $G \in \mathcal{G}_{0,1}^{m \times s}$ (and empirically for various random sparse and structured matrices G as well) the matrix $B = A^T G$ is expected (a) to have full rank and to be well conditioned if and only if $s \geq \rho$, (b) to approximate a matrix basis (within an error norm in $O(\sigma_{\rho+1}(A))$) for a linear space $\mathbb{T} \supseteq \mathbb{T}_{\rho,B} = \mathbb{T}_{\rho,A}$ where $s \geq \rho$, and (c) to approximate a matrix basis (within an error norm in $O(\sigma_{\rho+1}(A))$) for the space $\mathbb{T}_{\rho,A}$ where $s = \rho$. Property (a) is implied by Theorem 4.1, properties (b) and (c) by Theorem 7.2.

Proto-Algorithm 8.1. Numerical rank with random sampling (see Remarks 8.2–8.4).

INPUT: Two integers ρ_- and ρ_+ and a matrix $A \in \mathbb{R}^{m \times n}$ having unknown numerical rank $\rho = \text{rank}(A)$ in the range $[\rho_-, \rho_+]$ such that $0 \leq \rho_- < \rho_+ \leq n \leq m$, a rule for the selection of a candidate integer ρ in a range $[\rho_-, \rho_+]$, and a Subroutine COND that determines whether a given matrix has full rank and is well conditioned or not.

OUTPUT: an integer ρ expected to equal numerical rank of the matrix A and a matrix B expected to approximate (within an error norm in $O(\sigma_{\rho+1}(A))$) a matrix basis of the singular space $\mathbb{T}_{\rho,A}$. (Both expectations can actually fail, but with a low probability, see Remark 8.2.)

INITIALIZATION: Generate matrix $G \in \mathcal{G}_{0,1}^{m \times \rho_+}$ and write $B = A$, $G_\rho = G(I_\rho \mid O_{\rho, m-\rho})^T$ for $\rho = \rho_-, \rho_- + 1, \dots, \rho_+$.

COMPUTATIONS:

1. Output $\rho = \rho_+$ and the matrix B and stop if $\rho_- = \rho_+$. Otherwise fix an integer ρ in the range $[\rho_-, \rho_+]$.
2. Compute the matrix $B' = B^T G_\rho$ and apply to it the Subroutine COND.
3. If this matrix has full rank and is well conditioned, write $\rho_+ = \rho$ and $B = B'$ and go to Stage 1. Otherwise write $\rho_- = \rho$ and go to Stage 1.

Remark 8.2. The algorithm can output a wrong value of the numerical rank, although by virtue of Theorems 7.1 and 7.2 combined this occurs with a low probability. One can decrease this probability by reapplying the algorithm to the same inputs and choosing distinct random parameters. Furthermore one can fix a tolerance τ of order $\sigma_{\rho+1}(A)$, set $T = B$, and apply Stage 3 of Proto-Algorithm 7.1. Then $\text{nrnk}(A)$ is expected to exceed the computed value ρ if this stage outputs FAILURE and to equal ρ otherwise, in which case the algorithm also outputs a rank- ρ approximation of the matrix A (within an error norm $\tau \|A\|$ in $O(\sigma_{\rho+1}(A))$). For a sufficiently small tolerance τ the latter outcome implies that certainly $\rho \geq \text{nrnk}(A)$.

Remark 8.3. A Subroutine COND, which tests whether an $m \times \rho$ matrix B' has full rank and is well conditioned, can employ SVD of the matrix A or its rank revealing factorization, thus involving pivoting or orthogonalization. We can avoid this charge on matrix sparseness and structure by using randomization (although this is less important where ρ_+ is small). Namely assume that, say $m \geq n$ and recall that the algorithm of [D83] computes a close upper bound σ_+^2 on the largest eigenvalue σ^2 of the matrix $S = A^T A$ by recursively computing the vectors $\mathbf{v}_i = S^i \mathbf{v} = S \mathbf{v}_{i-1}$, $i = 1, 2, \dots$ for a random vector $\mathbf{v} = \mathbf{v}_0$. By reapplying this algorithm to the matrix $\sigma_+^2 I - A^T A$ we can approximate the absolutely smallest eigenvalue of the matrix S , which is actually equal to $\sigma_n^2(A)$. Here we just need a crude estimate to support our algorithm.

Remark 8.4. The binary search $\rho = \lceil (\rho_- + \rho_+)/2 \rceil$ is an attractive policy for choosing the candidate values ρ , but one may prefer to move toward the left end ρ_- of the range more rapidly, to decrease the size of the matrix B' .

In principle in our search for numerical rank we can employ Corollary 7.1 or 7.2 instead of Theorem 7.2. Then we would have to apply the Subroutine COND to matrices of size $m \times n$ or larger, which means extra computational cost. Because of that the two respective Prototype Algorithms below cannot compete with Proto-Algorithm 8.1 unless the input matrix has a small numerical nullity.

Proto-Algorithm 8.2. Numerical rank via randomized additive preprocessing.

INPUT, OUTPUT and Stage 1 of COMPUTATIONS are the same as in Proto-Algorithm 8.1.

INITIALIZATION: Compute the integer $r_+ = n - \rho_-$ and a scalar σ of order $\|A\|$, generate two matrices $U_+ \in \mathcal{G}_{0,\sigma}^{m \times r_+}$ and $V_+ \in \mathcal{G}_{0,\sigma}^{n \times r_+}$, and write $U_s = U_+(I_s \mid O_{s, m-s})^T$ and $V_s = V_+(I_s \mid O_{n-s, s})^T$ for $s = r_-, r_- + 1, \dots, r_+$.

COMPUTATIONS:

2. Compute the integer $s = n - \rho$. Compute the matrix $C = A + U_s V_s^T$ and apply to it the Subroutine COND.
3. If this matrix is rank deficient or ill conditioned write $\rho_+ = \rho$ and go to Stage 1. Otherwise write $\rho_- = \rho$ and go to Stage 1.

Proto-Algorithm 8.3. Numerical rank via randomized augmentation.

INPUT, OUTPUT and Stages 1 and 3 of COMPUTATIONS are the same as in Proto-Algorithm 8.2.

INITIALIZATION: Compute the integer $r_+ = n - \rho_-$ and a scalar σ of order $\|A\|$, generate three matrices $U_+ \in \mathcal{G}_{0,\sigma}^{m \times r_+}$, $V_+ \in \mathcal{G}_{0,\sigma}^{n \times r_+}$, and $W_+ \in \mathcal{G}_{0,\sigma}^{r_+ \times r_+}$, and write $i = 1$, $A_0 = A$, $U_s = U_+(I_s \mid O_{s,m-s})^T$, $V_s = V_+(I_s \mid O_{n-s,s})^T$, and $W_s = (I_s \mid O_{s,r_+-s})W_+(I_s \mid O_{s,r_+-s})^T$ for $s = r_-, r_- + 1, \dots, r_+$.

COMPUTATIONS:

2. Compute the integer $s = n - \rho$. Compute the matrix $K = \begin{pmatrix} W_s & V_s^T \\ -U_s & A \end{pmatrix}$ and apply to it the Subroutine COND.

8.3 Preprocessing for Newton–Toeplitz iteration

Newton’s iteration for matrix inversion

$$X_{i+1} = X_i(2I - CX_i), \quad i = 0, 1, \dots \quad (8.1)$$

squares the residuals $I - CX_i$, that is,

$$I - CX_{i+1} = (I - CX_i)^2 = (I - CX_0)^{2^{i+1}}, \quad i = 0, 1, \dots \quad (8.2)$$

Therefore

$$\|I - CX_{i+1}\| \leq \|I - CX_i\|^2 = \|I - CX_0\|^{2^{i+1}}, \quad i = 0, 1, \dots, \quad (8.3)$$

and so the approximations X_i quadratically converge to the inverse C^{-1} right from the start provided that $\|I - CX_0\| < 1$. We can ensure that $\|I - CX_0\| \leq 1 - \frac{2n}{(\kappa(C))^2(1+n)}$ by choosing $X_0 = \frac{2nC^T}{(1+n)\|C\|_1\|C\|_\infty}$ [PS91]. Newton’s iteration can be incorporated into our randomized algorithms. E.g., we can use it instead of Gaussian elimination in Proto-Algorithm 9.4 of the next section.

The map $C \Rightarrow X_0$ preserves the matrix structure of Toeplitz or Hankel type, but is the structure maintained throughout the iteration? Not automatically. In fact Newton’s loop can triple the displacement rank of a matrix X_k . The papers [P92], [P93], and [P93a], however, have proposed to maintain the structure via recursive compression of the displacements (one can also say *recompression*), thus defining *Newton’s structured* (e.g., Newton–Toeplitz) iteration. Recall that we can readily recover a Toeplitz-like matrix from its displacement (cf. (2.13)). According to the compression policy proposed in the papers [P92], [P93], and [P93a], one should periodically set to 0 the smallest singular values of the displacements of the matrices X_i to keep the length of the displacements within a fixed tolerance, equal to or a little exceeding the displacement rank of the input matrix C . According to the estimates in [P01], the Newton–Toeplitz iteration converges quadratically right from the start provided $\|I - CX_0\| < \frac{1}{(1+\|Z_e\|+\|Z_f\|)\kappa(C)}\|L^{-1}\|$, $\|L^{-1}\| \leq c_{e,f}n$, L denotes the associated displacement operator $L : C \Rightarrow Z_e C - C Z_f$ for $e \neq f$ or $L : C \Rightarrow C - Z_e C Z_f^T$ for $ef \neq 1$, and $c_{e,f}$ is a constant defined by e and f . Similar bounds can be deduced for other classes of matrices with displacement structure [P01, Section 6.6]. See [PBRZ99], [PRW02], [P01, Chapter 6] and [P10] on further information. The cost of computing the $n \times d$ generator matrices G and H with SVD or rank revealing factorization is not high for small ranks d , but randomized methods of Section 7 enable further cost decrease.

8.4 Application to tensor decomposition

Let

$$\mathbf{A} = [A(i_1, \dots, i_d)] \quad (8.4)$$

denote a d -dimensional *tensor* with entries $A(i_1, \dots, i_d)$ and *spacial indices* i_1, \dots, i_d ranging from 1 to n_1, \dots, n_d , respectively. Define the $d - 1$ *unfolding matrices* $A_k = [A(i_1 \dots i_k; i_{k+1} \dots i_d)]$, $k = 1, \dots, d$, where the semicolon separates the multi-indices $i_1 \dots i_k$ and $i_{k+1} \dots i_d$, which define the

rows and columns of the matrix A_k , respectively, $k = 1, \dots, d$. The paper [O09] proposed the following class of *Tensor Train Decompositions*, hereafter referred to as *TT Decompositions*, where the *summation indices* $\alpha_1, \dots, \alpha_{d-1}$ ranged from 1 to *compression ranks* r_1, \dots, r_{d-1} , respectively,

$$T = \sum_{\alpha_1, \dots, \alpha_{d-1}} G_1(i_1, \alpha_1) G_2(\alpha_1, i_1, \alpha_2) \cdots G_{d-1}(\alpha_{d-2}, i_{d-1}, \alpha_{d-1}) G_d(\alpha_d, i_d). \quad (8.5)$$

Theorem 8.1. [O09]. *For any tensor \mathbf{A} of (8.4) there exists a TT decomposition (8.5) such that $\mathbf{A} = \mathbf{T}$ and $r_k = \text{rank}(A_k)$ for $k = 1, \dots, d-1$.*

There is a large and growing number of important applications of TT decompositions (8.5) to modern computations (cf. e.g., [OT09], [OT10], [OT11]) where the numerical ranks of the unfolding matrices A_k are much smaller than their ranks, and it is desired to compress TT decompositions respectively.

Theorem 8.2. [OT10]. *For any tensor \mathbf{A} of (8.4) and any set of positive integers $r_k \leq \text{rank}(A_k)$, $k = 1, \dots, d-1$, there exists a TT decomposition (8.5) such that*

$$\|\mathbf{A} - \mathbf{T}\|_F^2 \leq \sum_{k=1}^{d-1} \tau_k^2, \quad \tau_k = \min_{\text{rank}(B)=r_k} \|A_k - B\|_F, \quad k = 1, \dots, d-1. \quad (8.6)$$

The constructive proof of this theorem in [OT10] relies on inductive approximation of unfolding matrices by their SVDs truncated to the compression ranks r_k . Let us sketch this construction. For $d = 2$ we obtain a desired TT decomposition $T(i_1, i_2) = \sum_{\alpha_1}^{r_1} G_1(i_1, \alpha_1) G_2(\alpha_1, i_2)$ (that is a sum of r_1 outer products of r_1 pairs of vectors) simply by truncating the SVD of the matrix $A(i_1, i_2)$. At the inductive step one truncates the SVD of the first unfolding matrix $A_1 = S_{A_1} \Sigma_{A_1} T_{A_1}^T$ to obtain rank- r_1 approximation of this matrix $B_1 = S_{B_1} \Sigma_{B_1} T_{B_1}^T$ where $\Sigma_{B_1} = \text{diag}(\sigma_j(A_1))_{j=1}^{r_1}$ and the matrices S_{B_1} and T_{B_1} are formed by the first r_1 columns of the matrices S_{A_1} and T_{A_1} , respectively. Then it remains to approximate the tensor $\mathbf{B} = [B(i_1, \dots, i_d)]$ represented by the matrix B_1 . Rewrite it as $\sum_{\alpha_1=1}^n S_{B_1}(i_1; \alpha_1) \hat{A}(\alpha_1; i_2 \dots i_d)$ for $\hat{A} = \sum_{B_1} T_{B_1}^T$, represent \hat{A} as the tensor $\hat{\mathbf{A}} = [A(\alpha_1 i_2, i_3, \dots, i_d)]$ of dimension $d-1$, apply the inductive hypothesis to obtain a TT-approximation of this tensor, and extend it to a TT-approximation of the original tensor \mathbf{A} .

In [OT10] the authors specify this construction as their Algorithm 1, prove error norm bound (8.6), then point out that the “computation of the truncated SVD for large scale and possibly dense unfolding matrices ... is unaffordable in many dimensions”, propose “to replace SVD by some other dyadic decompositions $A_k \approx UV^T$, which can be computed with low complexity”, and finally specify such recipe as [OT10, Algorithm 2], which is an iterative algorithm for skeleton or pseudoskeleton decomposition of matrices and which they use at Stages 5 and 6 of their Algorithm 1. The cost of each iteration of [OT10, Algorithm 2] is quite low and empirically the iteration converges fast, but the authors welcome alternative recipes having formal support.

Our randomized Proto-Algorithms 7.1 and 8.1 can serve as the alternatives to [OT10, Algorithm 2]. For the input matrix A_1 above they use $O(r_1)$ multiplications of this matrix by $O(r_1)$ vectors, which means a low computational cost for sparse and structured inputs, whereas the expected output is an approximate matrix basis for the space \mathbb{S}_{r_1, A_1} or \mathbb{T}_{r_1, A_1} and a rank- r_1 approximation to the matrix A_1 , within an expected error norm in $O(\sigma_{r_1+1}(A_1))$. This is the same order as in [OT10, Algorithm 1], but now we do not use SVDs. One can further decrease the error bound by means of small oversampling of Remark 7.2 and power transform of Remark 7.5.

9 2×2 block triangulation of an ill conditioned matrix, matrix inversion, and solving linear systems of equations

In this section we apply the results of Section 7 to compute 2×2 block triangulation and the inverse of a nonsingular ill conditioned matrix. In Section 9.3, which can be read independently of the rest of the present section, we combine additive preconditioning and the SMW formula to solve a linear

system of equations with such matrix. One can alternatively combine dual additive preprocessing with the dual SMW formula or apply these techniques to precondition the input matrix. (See Remarks 9.4 and 9.6). We partition some algorithms of this section into *symbolic* and *numerical* stages. At the former stage we perform computations with infinite precision, but they use a small fraction of the double precision flops involved.

9.1 Block triangulation using approximate trailing singular spaces

In Section 7.5 we have approximated leading and trailing singular spaces of ill conditioned matrices by applying randomized additive preprocessing, random sampling, or augmentation. Next we extend the former algorithms (based first on additive preprocessing of Theorem 7.5 and Corollary 7.1 and then on random sampling of Section 7.4) to 2×2 block triangulation of these matrices. One can similarly compute their block triangulation by extending either the algorithms of Section 7.5 based on randomized augmentation of Corollary 7.2 (we leave this to the reader) or the algorithms of Sections 7.4 and 7.6 for approximate matrix bases of the leading singular spaces of the input matrix (see the next subsection).

Proto-Algorithm 9.1. Block triangulation with randomization and orthogonalization.

INPUT: A matrix $A \in \mathbb{R}^{n \times n}$ whose norm $\|A\|$ is neither large nor small, its numerical rank q satisfying $0 < q = n - r < n$, and a Subroutine LIN·SOLVE that either solves a linear system of equations if it is nonsingular and well conditioned or outputs FAILURE otherwise.

OUTPUT: FAILURE (with a low probability) or four orthogonal matrices K_0 and L_0 in $\mathbb{R}^{n \times q}$ and K_1 and L_1 in $\mathbb{R}^{n \times r}$ such that with a probability near 1 the $q \times q$ block submatrix $W_{00} = K_0^T A L_0$ of the matrix $W = (K_0 \mid K_1)^T A (L_0 \mid L_1) = \begin{pmatrix} W_{00} & W_{01} \\ W_{10} & W_{11} \end{pmatrix}$ is nonsingular, well conditioned, and strongly dominant such that $\sigma_q(W_{00}) \gg \max\{\|W_{01}\|, \|W_{10}\|, \|W_{11}\|\}$.

COMPUTATIONS (see Remark 9.1):

1. Generate two matrices $U, V \in \mathbb{G}_{0,1}^{n \times r}$.
2. Compute the matrix $C = A + UV^T$, expected to be nonsingular and well conditioned.
3. Apply the Subroutine LIN·SOLVE to compute the matrices $C^{-T}V$ and $C^{-1}U$. Stop and output FAILURE if so does the subroutine.
4. Compute and output two orthogonal matrices $K_1 = Q(C^{-1}U)$ and $L_1 = Q(C^{-T}V)$.
5. Compute and output two orthogonal nmbs $K_0 = \text{nmbs}(K_1)$ and $L_0 = \text{nmbs}(L_1)$.

The algorithm can only fail with a low probability by virtue of Theorems 5.6 and 7.5 and Corollary 7.1. We use the following theorem to prove correctness of the algorithm.

Theorem 9.1. For a matrix $A \in \mathbb{R}^{m \times n}$ and $0 < q < l = \min\{n, m\}$, write $r = n - q$ and $\bar{r} = m - q$. Let $K_0 \in \mathbb{R}^{m \times q}$, $L_0 \in \mathbb{R}^{n \times q}$, $K_1 \in \mathbb{R}^{m \times \bar{r}}$, $L_1 \in \mathbb{R}^{n \times r}$, and $Q_K, Q_L \in \mathbb{R}^{r \times r}$ be six orthogonal matrices such that $K_1 = S_{A,\bar{r}} Q_K$, $L_1 = T_{A,r} Q_L$, $K_1^T K_0 = O_{\bar{r},q}$ and $L_1^T L_0 = O_{r,q}$. Then $\|K_1^T A\| \leq \sigma_{q+1}(A)$, $\|AL_1\| \leq \sigma_{q+1}(A)$, $\|K_0 A L_0\| = \sigma_1(A)$, and $\kappa(K_0 A L_0) = \sigma_1(A)/\sigma_q(A)$.

Proof. Suppose $A = S_A \Sigma_A T_A^T$ is SVD of (2.5). Then $AL_1 = S_A \Sigma_A T_A^T T_{A,r} Q_L = S_A \Sigma_A \begin{pmatrix} O_{n,q} \\ Q_L \end{pmatrix} = S_A \text{diag}(O_{m-r,n-r}, (\sigma_j(A))_{j=q+1}^n Q_L)$, and so $\|AL_1\| \leq \sigma_{q+1}(A)$ because S_A and Q_L are orthogonal matrices. Similarly obtain that $\|K_1^T A\| \leq \sigma_{q+1}(A)$.

Next deduce from the assumptions about L_0 and L_1 that $L_0 = T_{q,A} Q'_0$ for an orthogonal matrix $Q'_0 \in \mathbb{R}^{q \times q}$ and similarly that $K_0 = S_{q,A} Q_0$ for an orthogonal matrix $Q_0 \in \mathbb{R}^{q \times q}$. Therefore

$$K_0 A L_0 = Q'_0 S_{q,A}^T S_A \Sigma_A T_A^T T_{q,A} Q_0 = (Q'_0 \mid O_{m,\bar{r}}) \Sigma_A (Q_0 \mid O_{r,n})^T = Q'_0 \text{diag}(\sigma_j(A))_{j=1}^q Q_0,$$

and so $\|K_0 A L_0\| = \sigma_1(A)$, $\kappa(K_0 A L_0) = \sigma_1(A)/\sigma_q(A)$. \square

In Proto-Algorithm 9.1 we expect to have $\mathcal{R}(L_1) \approx \mathcal{T}_{A,r}$ by virtue of Theorem 7.5 and similarly to have $\mathcal{R}(K_1) \approx \mathcal{S}_{A,r}$. Theorem 9.1 implies that the norms $\|K_1^T A\|$ and $\|AL_1\|$ have an upper bound close to $\sigma_{q+1}(A)$, whereas $\kappa(K_0^T AL_0) \approx \sigma_1(A)/\sigma_q(A)$. Now correctness of the algorithm follows because the matrix A has numerical rank q .

We can proceed with nonorthogonal matrices K_0, K_1, L_0, L_1, Q_K , and Q_L to simplify the computations, by weakening numerical stability a little. Then we can still expect that the norms $\|W_{01}\|$, $\|W_{10}\|$, and $\|W_{11}\|$ have at most order $\sigma_{q+1}(A)$, the norm $\|W_{00}\|$ has order $\sigma_1(A)$, and the condition number $\kappa(W_{00})$ has order $\sigma_1(A)/\sigma_q(A)$. Moreover choosing random matrices $K_0 \in \mathcal{G}_{0,1}^{m \times q}$ and $L_0 \in \mathcal{G}_{0,1}^{n \times q}$, which are expected to be well conditioned by virtue of Theorems 3.1 and 3.2 combined, and we can still extend our probabilistic estimates for the values $\|W_{i,j}\|$ for $i, j = 1, 2$ and $\kappa(W_{00})$. Here is our respective simplified algorithm. Our tests in Section 10.4 show its efficiency.

Proto-Algorithm 9.2. Simplified randomized block triangulation.

INPUT, OUTPUT and Stages 1 and 2 of COMPUTATIONS are the same as in Proto-Algorithm 9.1 except that the output matrices K_0, L_0, K_1 and L_1 are no longer assumed to be orthogonal.

COMPUTATIONS:

3. Generate and output two random matrices $K_0, L_0 \in \mathcal{G}_{0,1}^{n \times q}$.
4. Apply the Subroutine LIN-SOLVE to compute and to output the matrices $K_1 = C^{-T}V$ and $L_1 = C^{-1}U$. Output FAILURE and stop if so does the subroutine.

Proto-Algorithms 9.1 and 9.2 do not produce block triangulation but prepare it. Having strong domination of the block W_{00} , we can readily compute the block factorizations

$$W = \begin{pmatrix} I & O \\ W_{10}W_{00}^{-1} & I \end{pmatrix} \begin{pmatrix} W_{00} & W_{01} \\ O & G \end{pmatrix}$$

for $G = W_{11} - W_{10}W_{00}^{-1}W_{01}$ and

$$W^{-1} = \begin{pmatrix} W_{00}^{-1} & -W_{00}^{-1}W_{01}G^{-1} \\ O & G^{-1} \end{pmatrix} \begin{pmatrix} I & O \\ -W_{10}W_{00}^{-1} & I \end{pmatrix}.$$

Both Proto-Algorithms 9.1 and 9.2 reduce the inversion of the matrix A to the inversion of the matrices W_{00} and G of smaller sizes, where both matrices are expected to be nonsingular and better conditioned than the matrix A (cf. [PGMQ, Section 9]).

Remark 9.1. We expect to arrive at the matrices W_{01}, W_{10} and W_{11} having small norms. To counter the expected cancellation of the leading digits of the $2rn - r^2$ entries of these matrices, we should compute the matrices C, K_1 and L_1 , their products by the blocks of the matrix A , and the Schur complement G with a high precision p_+ (or partly symbolically, with infinite precision). These computations involve $O(n^2r)$ flops, that is a r/n fraction of order n^3 flops in high precision p_+ required by Gaussian elimination. See further study in [PGMQ, Section 9]. Having implemented this part of the computations with higher precision, we have outperformed the standard algorithms (see Section 10.5 and Tables 10.11 and 10.12).

9.2 Block triangulation using approximate leading singular spaces

Suppose a square matrix A has a small positive numerical rank q and define a dual variation of Proto-Algorithm 9.1 by applying Theorem 7.6. In this case matrix inversions are limited to the case of $q \times q$ matrices $H, K_0^T K_0$ and $L_0^T L_0$. In our dual algorithm we assume that the nonsingular input matrix A has been scaled so that the norm $\|A^{-1}\|$ is neither large nor small. See some recipes for the approximation of this norm at the end of Section 5 and see Remark 9.3 on how to proceed without estimating this norm.

Proto-Algorithm 9.3. Block triangulation using approximate leading singular spaces.

INPUT: A nonsingular ill conditioned matrix $A \in \mathbb{R}^{n \times n}$ scaled so that the norm $\|A^{-1}\|$ is neither large nor small; the numerical rank q of the matrix A such that $0 < q = n - r < n$, and a Subroutine INVERT that either inverts a matrix if it is nonsingular and well conditioned or outputs FAILURE otherwise.

OUTPUT: FAILURE (with a low probability) or four matrices $K_0, L_0 \in \mathbb{R}^{n \times q}$ and $K_1, L_1 \in \mathbb{R}^{n \times r}$ such that

$$W = \begin{pmatrix} W_{00} & W_{01} \\ W_{10} & W_{11} \end{pmatrix} = (K_0 \mid K_1)^T A (L_0 \mid L_1)$$

and the block submatrix $W_{00} = K_0^T A L_0$ is expected to be nonsingular, well conditioned, and strongly dominant such that $\sigma_q(W_{00}) \gg \max\{\|W_{01}\|, \|W_{10}\|, \|W_{11}\|\}$.

COMPUTATIONS:

1. Generate two matrices U_- and V_- in $\mathcal{G}_{0,1}^{n \times q}$.
2. Compute the matrix $H = I_q + V_-^T A U_-$ of (2.9).
3. Apply the Subroutine INVERT to compute the matrix H^{-1} . Output FAILURE and stop if so does the subroutine.
4. Compute the matrix $C_- = A - A U_- H^{-1} V_-^T A$ of (2.10).
5. Compute and output the matrices $K_0 = C_- U_- / \|C_- U_-\|$ and $L_0 = C_-^T V_- / \|C_-^T V_-\|$.
6. Compute the matrices $M \approx \text{nmbs}(K_0^T)$ and $N \approx \text{nmbs}(L_0^T)$ (see our Section 7.5, [PQ10] and [PQa] on the approximation of nmbs).
7. Compute and output the matrices $K_1 = M / \|M\|$ and $L_1 = N / \|N\|$.

The algorithm fails with a low probability by virtue of Theorems 5.6 and 7.6 and Corollary 7.3. Complete the correctness proof by extending Theorem 9.1.

Remark 9.2. As in the previous subsection, we must perform a small fraction of our computations with high accuracy. Namely we must compute the matrix H with high or infinite precision, but for that we need $O(qn^2)$ flops, versus order n^3 high precision flops in Gaussian elimination. Unlike the previous subsection, this stage involves only matrix multiplications, a matrix addition and no inversions, although we need matrix inversions for computing nmbs at Stage 6.

Remark 9.3. Instead of applying Theorem 7.6 we can employ any other algorithm that computes a pair of approximate matrix bases K_0 and L_0 for the left and right leading singular spaces. E.g., we can apply randomized dual augmentation of Corollary 7.4 or just compute $K_0 = A^T V$ and $L_0 = A U$ for $V \in \mathcal{G}_{0,1}^{q \times n}$ and $U \in \mathcal{G}_{0,1}^{n \times q}$ (cf. Proto-Algorithm 7.1). In a heuristic variation we can choose the matrices $U, V^T \in \mathcal{T}_{0,1}^{n \times q}$ where A is a Toeplitz-like matrix and where the numerical rank q is not small. The latter computation requires no estimates for the norm $\|A^{-1}\|$ and in our tests has led to higher output accuracy than Proto-Algorithm 9.3. For a further heuristic simplification in the case where $n = 2q$, choose the Toeplitz matrices U and V in the form $F = (Z \mid T)$, $Z \in \mathcal{Z}_{0,1}^{q \times n}$ and $e_1^T \in \mathcal{Z}_{0,1}^{1 \times n}$, and then $\begin{pmatrix} Z^{-1} T \\ -I_q \end{pmatrix}$ is a $\text{nmbs}(F)$.

9.3 Randomized additive preconditioning with the SMW recovery

Suppose that we seek the solution $\mathbf{y} = A^{-1} \mathbf{b}$ of a real nonsingular ill conditioned linear system $A \mathbf{y} = \mathbf{b}$ of n equations where we are given a small upper bound r on the numerical nullity of A . Assume that the norm $\|A\|$ is neither large nor small. Then randomized additive preprocessing $A \Rightarrow C = A + UV^T$ for $U, V \in \mathcal{G}_{0,1}^{n \times r}$ is expected to produce a well conditioned matrix C (cf. Theorem 5.6). The SMW formula implies that $\mathbf{y} = C^{-1} \mathbf{b} + C^{-1} U G^{-1} V^T C^{-1} \mathbf{b}$ for $G = I_r - V^T C^{-1} U$. Substitute $X_U = C^{-1} U$ and $\mathbf{x}_b = C^{-1} \mathbf{b}$ and obtain

$$\mathbf{y} = \mathbf{x}_b + X_U G^{-1} V^T \mathbf{x}_b \text{ for } G = I_r - V^T X_U. \quad (9.1)$$

This reduces the computation of \mathbf{y} essentially to the solution of the matrix equation $CX = (U \mid \mathbf{b})$ for $X = (X_U \mid \mathbf{x}_b)$, computing the matrix G , and its inversion. The solution algorithm below incorporates iterative refinement at this stage.

Proto-Algorithm 9.4. Randomized Solution of a Linear System with Iterative Refinement.

INPUT: a vector $\mathbf{b} \in \mathbb{R}^{n \times 1}$, a nonsingular ill conditioned matrix $A \in \mathbb{R}^{n \times n}$, and its numerical nullity $r = \text{nnul}(A)$ (cf. Remark 9.5).

OUTPUT: A vector $\tilde{\mathbf{y}} \approx A^{-1}\mathbf{b}$.

COMPUTATIONS:

1. Generate two matrices $U, V \in \mathcal{G}_{0,\sigma}^{n \times r}$.
2. Compute the matrix $C = A + UV^T$, expected to be nonsingular and well conditioned.
3. Apply Gaussian elimination (or another direct algorithm) to compute an approximate inverse $Y \approx C^{-1}$. (Perform the computations in double precision. Application of the same algorithm to the original ill conditioned linear system $A\mathbf{y} = \mathbf{b}$ would require about as many flops but in extended precision.)
4. Apply iterative refinement employing the approximate inverse Y to compute sufficiently accurate solution X_U of the matrix equation $CX_U = U$. (High accuracy is required to counter the cancelation of leading bits in the subsequent computation of the Schur complement $G = I_r - V^T C^{-1} U$.) Then recover a close approximation to the vector $\mathbf{y} = A^{-1}\mathbf{b}$ by applying equation (9.1).

The algorithm reduces the original task of computations with ill conditioned matrix A to the computations with the well conditioned matrix C and $O(n^2 r)$ additional flops. To handle an ill conditioned input A , we must perform computations with extended precision to counter magnification of rounding errors, but we can confine this stage essentially to computing the Schur complement $G = I_r - V^T C^{-1} U$. This is a small fraction of the computational time of the customary algorithms for a linear system $A\mathbf{y} = \mathbf{b}$ provided the ratio r/n is small and the precision required to handle the ill conditioned matrix A is high.

Let us supply some estimates. To support iterative refinement we must use a precision p exceeding $\log_2 \kappa(C)$; for well conditioned matrices C we can assume that $p > 2 \log_2 \kappa(C)$, say. Then order $p - \log_2 \kappa(C)$ new correct bits are produced per an output value by a loop of iterative refinement (see [PGMQ, Section 9]), reduced essentially to multiplication of the matrices C and Y by $2r$ vectors, that is to $(4n - 2)nr$ flops in a low (e.g., double) precision p . The refinement algorithm outputs order rn values; one can accumulate them with high accuracy as the sums of sufficiently many low precision summands (as in symbolic lifting [P11]). Overall with this advanced implementation we only perform $O(rn^2 p_+/p)$ flops in low precision p at Stage 4 of Proto-Algorithm 9.4.

For comparison Gaussian elimination uses $\frac{2}{3}n^3 + O(n^2)$ flops in extended precision $p_+ \approx p_{\text{out}} + \log_2 \kappa(A)$ to output the solution to the ill conditioned linear system $A\mathbf{y} = \mathbf{b}$ with a prescribed precision p_{out} . We compute an approximate inverse Y of the well conditioned matrix C at Stage 3 by using $\frac{2}{3}n^3 + O(n^2)$ flops as well, but in the low precision p . The cost of performing Stages 1 and 2 is dominated, and so our progress is significant where $np \gg rp_+$ and p_+ greatly exceeds p .

The estimated computational cost further decreases where the matrices A , U and V have consistent patterns of sparseness and structure. E.g., the decrease is by a factor n where they are Toeplitz or Toeplitz-like matrices.

Remark 9.4. Given a nonsingular $n \times n$ matrix A (with $\|A^{-1}\| \approx 1$) and a small upper bound q on its numerical rank, we can define a dual variation of Proto-Algorithm 9.4 based on Corollary 7.3 as follows: generate a pair of matrices $U_-, V_- \in \mathbb{G}_{0,1}^{n \times q}$ and then compute the matrices H of (2.9)

and C_- of (2.10) to reduce the solution of a linear system of equations $A\mathbf{y} = \mathbf{b}$ to computing the vector $\mathbf{y} = (C_-^{-1} - U_-V_-^T)\mathbf{b}$ (cf. (2.9)). The matrix $H = I_q + V_-^T A U_-$ must be computed with high accuracy, but this only requires $O(qn^2)$ flops in high or infinite precision. Furthermore, unlike the computations by means of the SMW formula, we only need matrix multiplications and an addition (and no inversions) at this stage. Similarly one can employ Corollary 7.4 and expression (7.4) instead of Corollary 7.3.

Remark 9.5. There is no point for applying Proto-Algorithm 9.4 where the matrix A is well conditioned or has numerical nullity exceeding r . In the former case the preconditioning is not needed, whereas in the latter case additive preprocessing would produce an ill conditioned matrix C . In both cases preprocessing would give no benefits but would involve extraneous computations and additional rounding errors. In the case where r is equal to the numerical nullity of A , however, these deficiencies can be overwhelmed by the benefits of avoiding order n^3 high precision flops.

Remark 9.6. Instead of using additive preconditioners directly for solving linear systems, one can combine them with the SMW or dual SMW formulae to obtain multiplicative preconditioners. Assume a nonsingular ill conditioned $n \times n$ matrix A and let A_- denote the inverse A^{-1} computed with double precision. Then the matrices A_-A and AA_- are much better conditioned than the matrix A according to the experiments of [R90]. Note that both linear systems of equations $A_-A\mathbf{y} = A_-\mathbf{b}$ and $AA_-\mathbf{x} = \mathbf{b}$, for $\mathbf{y} = A_-\mathbf{x}$, are equivalent to the system $A\mathbf{y} = \mathbf{b}$. This empirical technique is interesting itself and can probably be advanced by means of its recursive application. It may also accentuate the preconditioning power of our randomized preprocessing. Instead of defining the multiplier A_- as the inverse A^{-1} computed with double precision, we can compute this multiplier as $C^{-1}(I_n - UG^{-1}V^TC^{-1})$ by applying the SMW formula. Moreover we can drop the factor C^{-1} and write either $A_- = I_n - UG^{-1}V^TC^{-1}$ or $A_- = I_n - C^{-1}UG^{-1}V^T$ to have $AA_- = A - AUG^{-1}V^TC^{-1}$ and $A_-A = A - UG^{-1}V^TC^{-1}A$ or $AA_- = A - AC^{-1}UG^{-1}V^T$ and $A_-A = A - C^{-1}UG^{-1}V^TA$. We can similarly utilize the dual SMW formula of (2.9) and (2.10), in which case we can compute the multiplier $A_- = C_-^{-1} - U_-V_-^T$ for C_- of (2.10), and then we would have $AA_- = (I - AUH^{-1}V^T)^{-1} - AU_-V_-^T$ and $A_-A = (I - UH^{-1}V^TA)^{-1} - U_-V_-^TA$ for $H = I_q + V_-^T U_-T$ of (2.9).

10 Numerical Experiments

Our numerical experiments with random general, Hankel, Toeplitz and circulant matrices have been performed in the Graduate Center of the City University of New York on a Dell server with a dual core 1.86 GHz Xeon processor and 2G memory running Windows Server 2003 R2. The test Fortran code was compiled with the GNU gfortran compiler within the Cygwin environment. Random numbers were generated with the random_number intrinsic Fortran function, assuming the uniform probability distribution over the range $\{x : -1 \leq x < 1\}$. The tests have been designed by the first author and performed by his coauthors.

10.1 Conditioning tests

We have computed the condition numbers of random general $n \times n$ matrices for $n = 2^k$, $k = 5, 6, \dots$, with entries sampled in the range $[-1, 1)$ as well as complex general, Toeplitz, and circulant matrices whose entries had real and imaginary parts sampled at random in the same range $[-1, 1)$. We performed 100 tests for each class of inputs, each dimension n , and each nullity r . Tables 10.2–10.4 display the test results. The last four columns of each table display the average (mean), minimum, maximum, and standard deviation of the computed condition numbers of the input matrices, respectively. Namely we computed the values $\kappa(A) = \|A\| \|A^{-1}\|$ for general, Toeplitz, and circulant matrices A and the values $\kappa_1(A) = \|A\|_1 \|A^{-1}\|_1$ for Toeplitz matrices A . We computed and displayed in Table 10.3 the 1-norms of Toeplitz matrices and their inverses rather than their 2-norms to facilitate the computations in the case of inputs of large sizes. Relationships (2.2) link the 1-norms and 2-norms to one another, but the empirical data in Table 10.1 consistently show even closer links, in all cases of general, Toeplitz, and circulant $n \times n$ matrices A where $n = 32, 64, \dots, 1024$.

10.2 Preconditioning tests

Table 10.5 covers our tests for the preconditioning power of additive preprocessing in [PIMR10]. We have tested the input matrices of the following classes.

1n. *Nonsymmetric matrices of type I with numerical nullity r .* $A = S\Sigma_r T^T$ are $n \times n$ matrices where S and T are $n \times n$ random orthogonal matrices, that is, the factors Q in the QR factorizations of random real matrices; $\Sigma_r = \text{diag}(\sigma_j)_{j=1}^n$ is the diagonal matrix such that $\sigma_{j+1} \leq \sigma_j$ for $j = 1, \dots, n-1$, $\sigma_1 = 1$, the values $\sigma_2, \dots, \sigma_{n-r-1}$ are randomly sampled in the semi-open interval $[0.1, 1)$, $\sigma_{n-r} = 0.1$, $\sigma_j = 10^{-16}$ for $j = n-r+1, \dots, n$, and therefore $\kappa(A) = 10^{16}$ [H02, Section 28.3].

1s. *Symmetric matrices of type I with numerical nullity r .* The same as in part 1n, but for $S = T$.

The matrices of the six other classes have been constructed in the form of $\frac{A}{\|A\|} + \beta I$ where the recipes for defining the matrices A and scalars β are specified below.

2n. *Nonsymmetric matrices of type II with numerical nullity r .* $A = (W \mid WZ)$ where W and Z are random orthogonal matrices of sizes $n \times (n-r)$ and $(n-r) \times r$, respectively.

2s. *Symmetric matrices of type II with numerical nullity r .* $A = WW^T$ where W are random orthogonal matrices of size $n \times (n-r)$.

3n. *Nonsymmetric Toeplitz-like matrices with numerical nullity r .* $A = c(T \mid TS)$ for random Toeplitz matrices T of size $n \times (n-r)$ and S of size $(n-r) \times r$ and for a positive scalar c such that $\|A\| \approx 1$.

3s. *Symmetric Toeplitz-like matrices with numerical nullity r .* $A = cTT^T$ for random Toeplitz matrices T of size $n \times (n-r)$ and a positive scalar c such that $\|A\| \approx 1$.

4n. *Nonsymmetric Toeplitz matrices with numerical nullity 1.* $A = (a_{i,j})_{i,j=1}^n$ is a Toeplitz $n \times n$ matrix. Its entries $a_{i,j} = a_{i-j}$ are random for $i-j < n-1$. The entry $a_{n,1}$ is selected to ensure that the last row is linearly expressed through the other rows.

4s. *Symmetric Toeplitz matrices with numerical nullity 1.* $A = (a_{i,j})_{i,j=1}^n$ is a Toeplitz $n \times n$ matrix. Its entries $a_{i,j} = a_{i-j}$ are random for $|i-j| < n-1$, whereas the entry $a_{1,n} = a_{n,1}$ is a root of the quadratic equation $\det A = 0$. We repeatedly generated the matrices A until we arrived at the quadratic equation having real roots.

We set $\beta = 10^{-16}$ for symmetric matrices A in the classes 2s, 3s, and 4s, so that $\kappa(A) = 10^{16} + 1$ in these cases. For nonsymmetric matrices A we defined the scalar β by an iterative process such that $\|A\| \approx 1$ and $10^{-18}\|A\| \leq \kappa(A) \leq 10^{-16}\|A\|$ [PIMR10, Section 8.2]. Table 10.5 displays the average values of the condition numbers $\kappa(C)$ of the matrices $C = A + UU^T$ over 100,000 tests for the inputs in the above classes, $r = 1, 2, 4, 8$ and $n = 100$. We defined the additive preprocessor UU^T by a normalized $n \times r$ matrix $U = U/\|U\|$ where $U^T = (\pm I \mid O_{r,r} \mid \pm I \mid O_{r,r} \mid \dots \mid O_{r,r} \mid \pm I \mid O_{r,s})$, we chosen the integer s to obtain $n \times r$ matrices U and chosen the signs for the matrices $\pm I$ at random. In our further tests the condition numbers of the matrices $C = A + 10^p UV^T$ for $p = -10, -5, 5, 10$ were steadily growing within a factor $10^{|p|}$ as the value $|p|$ was growing. This showed the importance of proper scaling of the additive preprocessor UV^T .

10.3 GENP with random circulant multipliers

Table 10.6 shows the results of our tests of the solution of a nonsingular well conditioned linear system $A\mathbf{y} = \mathbf{b}$ of n equations whose coefficient matrix had ill conditioned $n/2 \times n/2$ leading principal block for $n = 64, 256, 1024$. We performed 100 numerical tests for each dimension n and computed the maximum, minimum and average relative residual norms $\|A\mathbf{y} - \mathbf{b}\|/\|\mathbf{b}\|$ as well as standard deviation. GENP applied to these systems has output corrupted solutions with residual norms ranging from 10 to 10^8 . When we preprocessed the systems with circulant multipliers filled with ± 1 (choosing the n signs \pm at random), the norms decreased to at worst 10^{-7} for all inputs. Table 10.6 also shows further decrease of the norm in a single step of iterative refinement. Table 2 in [PQZa] shows similar results of the tests where the input matrices were chosen similarly but so that all their leading blocks had numerical nullities 0 and 1 and where Householder multipliers

$I_n - \mathbf{u}\mathbf{v}^T / \mathbf{u}^T \mathbf{v}$ replaced the circulant multipliers. Here \mathbf{u} and \mathbf{v} denote two vectors filled with integers 1 and -1 under random choice of the signs $+$ and $-$.

10.4 Approximation of the tails and heads of SVDs and low-rank approximation of a matrix

At some specified stages of our tests of this subsection and Section 10.5 we performed additions, subtractions and multiplications with infinite precision (hereafter referred to as *error-free ring operations*). At the other stages we performed computations with double precision, and we rounded to double precision all random values. We performed two refinement iterations for the computed solution of every linear system of equations and matrix inverse.

Table 10.7 shows the data from our tests on the approximation of trailing singular spaces of the SVD of an $n \times n$ matrix A having numerical nullity $r = n - q$ and on the approximation of this matrix with a matrix of rank $q = n - r$. For $n = 64, 128, 256$ and $q = 1, 8, 32$ we generated $n \times n$ random orthogonal matrices S and T and diagonal matrices $\Sigma = \text{diag}(\sigma_j)_{j=1}^n$ such that $\sigma_j = 1/j$, $j = 1, \dots, q$, $\sigma_j = 10^{-10}$, $j = q + 1, \dots, n$ (cf. [H02, Section 28.3]). Then we applied error-free ring operations to compute the input matrices $A = S_A \Sigma_A T_A^T$, for which $\|A\| = 1$ and $\kappa(A) = 10^{10}$. Furthermore we generated pairs of random $n \times r$ matrices U and V for $r = 1, 8, 32$, scaled them to $\|UV^T\| \approx 1$, and computed the matrices $C = A + UV^T$ (by applying error-free ring operations), $B_{A,r} = C^{-1}U$ (by using two refinement iterations), $T_{A,r}$, $B_{A,r}Y_{A,r}$ as a least-squares approximation to $T_{A,r}$, $Q = Q(B_{A,r})$, and $A - AQQ^T$ (by applying error-free ring operations). Table 10.7 summarizes the data on the values $\kappa(C)$ and the residual norms $\text{rn}_1 = \|B_{A,r}Y_{A,r} - T_{A,r}\|$ and $\text{rn}_2 = \|A - AQQ^T\|$ observed in 100 runs of our tests for every pair of n and q .

We performed similar tests on the approximation of leading singular spaces of the SVDs of the same $n \times n$ matrices A having numerical rank q and numerical nullity $r = n - q$ and on the approximation of this matrix with a matrix of rank q . In some tests we employed dual additive preprocessing to approximate matrix bases for the leading singular spaces $\mathbb{T}_{q,A}$. We have generated the pairs of $n \times q$ random matrices U_- and V_- for $q = 1, 8, 32$, scaled them to have $\|U_- V_-^T\| \approx \|A^{-1}\| = 10^{10}$, and successively computed the matrices $H = I_q + V_-^T A U_-$ of (2.9) (by applying error-free ring operations), H^{-1} (by using two refinement iterations), $C_- = A - A U_- H^{-1} V_-^T A$ of (2.10), $B_{q,A} = C_-^T V_-$, $T_{q,A}$, $B_{q,A}Y_{q,A}$ as a least-squares approximation to $T_{q,A}$, $Q_{q,A} = Q(B_{q,A})$, and $A - A Q_{q,A} (Q_{q,A})^T$ (by applying error-free ring operations). Table 10.8 summarizes the data on the condition numbers $\kappa(C_-)$ and the residual norms $\text{rn}^{(1)} = \|B_{q,A}Y_{q,A} - T_{q,A}\|$ and $\text{rn}^{(2)} = \|A - A Q_{q,A} (Q_{q,A})^T\|$ obtained in 100 runs of our tests for every pair of n and q .

We have also performed similar tests where we generated random $n \times q$ matrices U (for $q = 1, 8, 32$) and random Toeplitz $n \times q$ matrices \bar{U} (for $q = 8, 32$) and then replaced the above matrix $B_{q,A}$ with the approximate matrix bases $A^T U$ and $A^T \bar{U}$ for the leading singular space $\mathbb{T}_{q,A}$. Tables 10.9 and 10.10 display the results of these tests. In both cases the residual norms are equally small and are about as small as in Tables 10.7 and 10.8.

10.5 2×2 block factorization and solving linear systems of equations

For our next tests we have chosen $n = 32, 64$ and $r = 1, 2, 4$ and for every pair $\{n, r\}$ generated 100 instances of vectors \mathbf{b} and matrices A , U , and V as follows. We generated (a) random vectors \mathbf{b} of dimension n , (b) random real orthogonal $n \times n$ matrices S and T , and (c) random $n \times r$ matrices U and V , which we scaled to have $\|U\| = \|V\| = 1$. Then we defined the matrices $\Sigma = \text{diag}(\sigma_j)_{j=1}^n$, with $\sigma_{n-j} = 10^{-17}$ for $j = 0, 1, \dots, r - 1$, and $\sigma_{n-j} = 1/(n - j)$ for $j = r, \dots, n - 1$, and then applied error-free ring operations to compute the matrices $A = S \Sigma T^T$. Note that $\|A\| = 1$ and $\|A^{-1}\| = 10^{17}$.

For every such input we solved the linear systems $A\mathbf{y} = \mathbf{b}$ by applying Proto-Algorithm 9.1. We first generated random $n \times (n - r)$ matrices K_0 and L_0 and then computed the matrices $C = A + UV^T$ (by applying error-free ring operations), $K_1 = C^{-T}V$ and $L_1 = C^{-1}U$ (by using two refinement iterations), and $W = (K_0 \mid K_1)^T A (L_0 \mid L_1) = \begin{pmatrix} W_{00} & W_{01} \\ W_{10} & W_{11} \end{pmatrix}$ (by applying error-free ring operations).

In all our tests the matrices C were nonsingular and well conditioned, and the leading principal $(n - r) \times (n - r)$ blocks $W_{00} = K_0^T A L_0$ were well conditioned and strongly dominated the three other blocks W_{01} , W_{10} , and W_{11} in the 2×2 block matrices W , as we expected to see by virtue of our analysis in Section 9.1. Then we computed the vector $\hat{\mathbf{b}} = (K_0 \mid K_1)^T \mathbf{b}$ (by applying error-free ring operations) and solved the linear system $W\mathbf{x} = \hat{\mathbf{b}}$ (by using two refinement iterations). Finally we computed the vector $\mathbf{y} = (L_0 \mid L_1)\mathbf{x}$ (by applying error-free ring operations). Table 10.11 shows the average (mean) values of the relative residual norms $\|\mathbf{A}\mathbf{y} - \mathbf{b}\|/\|\mathbf{b}\|$ of the output vectors \mathbf{y} (these values range about 10^{-10}) as well as the minimums, maximums, and standard deviations in these tests. For the same ill conditioned inputs the Subroutine MLDIVIDE(A,B) for Gaussian elimination from MATLAB has produced corrupted outputs, as can be seen from Table 10.12.

We have also performed similar tests for $n = 32, 64$ and $n \times n$ matrices A and vectors \mathbf{b} . We generated them as before, but for $q = n - r = 1, 2, 4$, and then we computed orthogonal matrices K_0 , K_1 , L_0 and L_1 by employing dual additive preprocessing and Proto-Algorithm 9.3. We first generated and scaled the pairs of random $n \times q$ matrices U_- and V_- such that $\|U_-\| \approx \|V_-\| \approx 3 \cdot 10^8$, and so $\|U_-\| \|V_-\| \approx \|A^{-1}\| = 10^{17}$. Then we successively computed the matrices H (by applying error-free ring operations), H^{-1} (by using two refinement iterations), $C_- = A - AU_-H^{-1}V_-^T A$ of (2.10), C_-U_- and $C_-^T V_-$ (all of them by applying error-free ring operations), $K_0 = Q(C_-U_-)$, $L_0 = Q(C_-^T V_-)$, $K_1 = Q(\text{numb}(K_0^T))$ and $L_1 = Q(\text{numb}(L_0^T))$, and continued as in the tests for Table 10.11. We displayed the results in Table 10.13, showing the residual norms of the order 10^{-9} on the average.

Furthermore we have performed similar tests where we first generated random $n \times q$ matrices U and V and then computed the matrix products $A^T V$ and AU (by applying error-free ring operations), and replaced the above matrices K_0 and L_0 by $K_0 = Q(A^T V)$ and $L_0 = Q(AU)$. Table 10.14 displays the results of these tests, showing the residual norms of order 10^{-25} on the average. Then again for the same ill conditioned inputs the Subroutine MLDIVIDE(A,B) for Gaussian elimination from MATLAB produced corrupted outputs, similarly to the results in Table 10.12.

We applied Proto-Algorithm 9.4 to solve linear systems of equations with the same inputs as above for small integers r . In these computations we used two refinement iterations for computing an approximate inverse $Y \approx G^{-1}$ and the solutions X_U and \mathbf{x}_b to the $r+1$ linear systems of equations $CX_U = U$ and $C\mathbf{x}_b = \mathbf{b}$, all with the matrix C . We computed the following matrices and vectors by applying error-free ring operations, $C = A + UV^T$, $G = I_r - V^T X_U$, and $\mathbf{y} = \mathbf{x}_b + X_U Y V^T \mathbf{x}_b$ of (9.1). Table 10.15 shows the test results for the same inputs as we used for tests of Table 10.11, except that now we have doubled the matrix size to $n = 64$ and $n = 128$.

10.6 Solution of a real symmetric Toeplitz linear system of equations with randomized augmentation

We have solved 100 real symmetric linear systems of equations $T\mathbf{y} = \mathbf{b}$ for each n where we used vectors \mathbf{b} with random coordinates from the range $[-1, 1)$ and Toeplitz matrices $T = S + 10^{-9}I_n$ for a singular symmetric Toeplitz $n \times n$ matrices S having rank $n - 1$ and nullity 1 and generated according to the recipe in [PQ10, Section 10.1b]. Table 10.16 shows the average CPU time of the solutions by our Algorithm 6.1 and, for comparison, based on the QR factorization and SVD, which we computed by applying the LAPACK procedures DGEQRF and DGESVD, respectively. To solve the auxiliary Toeplitz linear system $K\mathbf{x} = \mathbf{e}_1$ in Algorithm 6.1, we first employed the Toeplitz linear solver of [KV99], [V99], [VBHK01], and [VK98] and then applied iterative refinement with double precision.

The abbreviations “Alg. 6.1”, “QR”, and “SVD” indicate the respective algorithms. The last two columns of the table display the ratios of these data in the first and the two other columns. We measured the CPU time with the mclock function by counting cycles. One can convert them into seconds by dividing their number by a constant CLOCKS_PER_SEC, which is 1000 on our platform. We marked the table entries by a “-” where the tests were running too long and have not been completed. We obtained the solutions \mathbf{y} with the relative residual norms of about 10^{-15} in all three algorithms, which showed that Algorithm 6.1 employing iterative refinement was as reliable as the QR and SVD based solutions but ran much faster. We refer the reader to [PQZC, Table 3]

on similar test results for the solution of ill conditioned homogeneous Toeplitz linear systems.

Table 10.1: The norms of random general, Toeplitz and circulant matrices and of their inverses

| matrix A | n | $\ A\ _1$ | $\ A\ _2$ | $\frac{\ A\ _1}{\ A\ _2}$ | $\ A^{-1}\ _1$ | $\ A^{-1}\ _2$ | $\frac{\ A^{-1}\ _1}{\ A^{-1}\ _2}$ |
|------------------------------|------|-------------------|-------------------|---------------------------|-------------------|-------------------|-------------------------------------|
| General | 32 | 1.9×10^1 | 1.8×10^1 | 1.0×10^0 | 4.0×10^2 | 2.1×10^2 | 1.9×10^0 |
| General | 64 | 3.7×10^1 | 3.7×10^1 | 1.0×10^0 | 1.2×10^2 | 6.2×10^1 | 2.0×10^0 |
| General | 128 | 7.2×10^1 | 7.4×10^1 | 9.8×10^{-1} | 3.7×10^2 | 1.8×10^2 | 2.1×10^0 |
| General | 256 | 1.4×10^2 | 1.5×10^2 | 9.5×10^{-1} | 5.4×10^2 | 2.5×10^2 | 2.2×10^0 |
| General | 512 | 2.8×10^2 | 3.0×10^2 | 9.3×10^{-1} | 1.0×10^3 | 4.1×10^2 | 2.5×10^0 |
| General | 1024 | 5.4×10^2 | 5.9×10^2 | 9.2×10^{-1} | 1.1×10^3 | 4.0×10^2 | 2.7×10^0 |
| Toeplitz | 32 | 1.8×10^1 | 1.9×10^1 | 9.5×10^{-1} | 2.2×10^1 | 1.3×10^1 | 1.7×10^0 |
| Toeplitz | 64 | 3.4×10^1 | 3.7×10^1 | 9.3×10^{-1} | 4.6×10^1 | 2.4×10^1 | 2.0×10^0 |
| Toeplitz | 128 | 6.8×10^1 | 7.4×10^1 | 9.1×10^{-1} | 1.0×10^2 | 4.6×10^1 | 2.2×10^0 |
| Toeplitz | 256 | 1.3×10^2 | 1.5×10^2 | 9.0×10^{-1} | 5.7×10^2 | 2.5×10^2 | 2.3×10^0 |
| Toeplitz | 512 | 2.6×10^2 | 3.0×10^2 | 8.9×10^{-1} | 6.9×10^2 | 2.6×10^2 | 2.6×10^0 |
| Toeplitz | 1024 | 5.2×10^2 | 5.9×10^2 | 8.8×10^{-1} | 3.4×10^2 | 1.4×10^2 | 2.4×10^0 |
| Circulant | 32 | 1.6×10^1 | 1.8×10^1 | 8.7×10^{-1} | 9.3×10^0 | 1.0×10^1 | 9.2×10^{-1} |
| Circulant | 64 | 3.2×10^1 | 3.7×10^1 | 8.7×10^{-1} | 5.8×10^0 | 6.8×10^0 | 8.6×10^{-1} |
| Circulant | 128 | 6.4×10^1 | 7.4×10^1 | 8.6×10^{-1} | 4.9×10^0 | 5.7×10^0 | 8.5×10^{-1} |
| Circulant | 256 | 1.3×10^2 | 1.5×10^2 | 8.7×10^{-1} | 4.7×10^0 | 5.6×10^0 | 8.4×10^{-1} |
| Circulant | 512 | 2.6×10^2 | 3.0×10^2 | 8.7×10^{-1} | 4.5×10^0 | 5.4×10^0 | 8.3×10^{-1} |
| Circulant | 1024 | 5.1×10^2 | 5.9×10^2 | 8.7×10^{-1} | 5.5×10^0 | 6.6×10^0 | 8.3×10^{-1} |

11 Related work, our progress, and further study

Preconditioned iterative algorithms for linear systems of equations is a classical subject [A94], [B02], [G97]. The problem of creating inexpensive preconditioners for general use has been around for a long while as well. On estimating the condition numbers of random matrices see [D88], [E88], [ES05], [CD05], [SST06], [B11], and the bibliography therein. The study in the case of random structured matrices was stated as a challenge in [SST06]. We provide such estimates for Gaussian random Toeplitz and circulant matrices in Sections 3.4 and 3.5. They can be surprising because the paper [BG05] has proved that the condition numbers grow exponentially in n as $n \rightarrow \infty$ for some large and important classes of $n \times n$ Toeplitz matrices, whereas we prove the opposit for Gaussian random Toeplitz $n \times n$ matrices.

Our present study of randomized preconditioning extends [PGMQ], [PIMR10], [PQZa], and [PQZC]. In Sections 7–10 we outline and test some new applications of randomized preconditioning, whereas our Theorems 4.1 and 5.6 and and Corollaries 6.1 and 6.2 support such applications formally. The formal support relies on using Gaussian random matrices, but empirically our algorithms remained as efficient where instead we employed Gaussian random Toeplitz matrices and quite typically where we further decreased the number of random parameters involved, e.g., where we used block vectors U with the blocks $\pm I$ and O in the map $A \Rightarrow C = A + UU^T$ (see the end of Section 10.2 and Table 10.5) and used circulant or Householder multipliers filled with ± 1 (see our Section 10.3 and Table 10.6 and [PQZa, Table 2]), thus limiting randomization to the choice of the block sizes in the block vector U and of the signs \pm for the blocks $\pm I$ and entries ± 1 .

Besides the cited estimates for the condition numbers of Gaussian random Toeplitz and circulant matrices, our technical novelties include randomized multiplicative and additive preconditioning, dual additive preprocessing, augmentation and dual augmentation techniques, extensions of the SMW formula, proof of the power of all these techniques, randomized algorithms for numerical rank and approximation of leading and trailing singular spaces avoiding orthogonalization and pivoting and their applications to low-rank approximation of a matrix, tensor decomposition, and 2×2 block

Table 10.2: The condition numbers $\kappa(A)$ of random matrices A

| n | input | min | max | mean | std |
|------|---------|-------------------|-------------------|-------------------|-------------------|
| 32 | real | 2.4×10^1 | 1.8×10^3 | 2.4×10^2 | 3.3×10^2 |
| 32 | complex | 2.7×10^1 | 8.7×10^2 | 1.1×10^2 | 1.1×10^2 |
| 64 | real | 4.6×10^1 | 1.1×10^4 | 5.0×10^2 | 1.1×10^3 |
| 64 | complex | 5.2×10^1 | 4.2×10^3 | 2.7×10^2 | 4.6×10^2 |
| 128 | real | 1.0×10^2 | 2.7×10^4 | 1.1×10^3 | 3.0×10^3 |
| 128 | complex | 1.3×10^2 | 2.5×10^3 | 3.9×10^2 | 3.3×10^2 |
| 256 | real | 2.4×10^2 | 8.4×10^4 | 3.7×10^3 | 9.7×10^3 |
| 256 | complex | 2.5×10^2 | 1.4×10^4 | 1.0×10^3 | 1.5×10^3 |
| 512 | real | 3.9×10^2 | 7.4×10^5 | 1.8×10^4 | 8.5×10^4 |
| 512 | complex | 5.7×10^2 | 3.2×10^4 | 2.3×10^3 | 3.5×10^3 |
| 1024 | real | 8.8×10^2 | 2.3×10^5 | 8.8×10^3 | 2.4×10^4 |
| 1024 | complex | 7.2×10^2 | 1.3×10^5 | 5.4×10^3 | 1.4×10^4 |
| 2048 | real | 2.1×10^3 | 2.0×10^5 | 1.8×10^4 | 3.2×10^4 |
| 2048 | complex | 2.3×10^3 | 5.7×10^4 | 6.7×10^3 | 7.2×10^3 |

Table 10.3: The condition numbers $\kappa_1(A) = \frac{\|A\|_1}{\|A^{-1}\|_1}$ of random Toeplitz matrices A

| n | min | mean | max | std |
|------|-------------------|-------------------|-------------------|-------------------|
| 256 | 9.1×10^2 | 9.2×10^3 | 1.3×10^5 | 1.8×10^4 |
| 512 | 2.3×10^3 | 3.0×10^4 | 2.4×10^5 | 4.9×10^4 |
| 1024 | 5.6×10^3 | 7.0×10^4 | 1.8×10^6 | 2.0×10^5 |
| 2048 | 1.7×10^4 | 1.8×10^5 | 4.2×10^6 | 5.4×10^5 |
| 4096 | 4.3×10^4 | 2.7×10^5 | 1.9×10^6 | 3.4×10^5 |
| 8192 | 8.8×10^4 | 1.2×10^6 | 1.3×10^7 | 2.2×10^6 |

factorizations of ill conditioned matrices.

Approximation by low-rank matrices and the extensions to tensor decompositions are thriving research areas, with numerous applications to matrix and tensor computations. One can partly trace its previous study through the papers [GTZ97], [GT01], [GOS08], [HMT11], [KB09], [MMD08], [OT09], [OT10], [OT11], [T00], and the bibliography therein, but much earlier advances in this area appeared in the papers [BCLR79], [B80], [B85], [B86], [BC87], directed to estimating the border rank of matrices and tensors and initially motivated by the design of fast matrix multiplication algorithms. Presently, application of tensor decompositions to the acceleration of matrix computations is a fashionable subject with applications to many important areas of modern computing (see, e.g., [T00], [MMD08], [OT09], [KB09]), but then again its earliest examples appeared in the cited papers on border rank and in [P72]. The latter paper has introduced the technique of trilinear aggregation, a basic ingredient of all subsequent fast algorithms for matrix multiplication with the inputs of both immense sizes (far beyond any practical interest) [P79], [P81], [P84], [CW90], [S10], [VW12] and realistic moderate sizes [P81], [P84], [LPS92], [K04]. Historically this work was the first example of the acceleration of fundamental matrix computations by means of tensor decomposition. More special direction of Tensor Train (TT) decomposition was proposed in [O09] and further developed in [OT09], [OT10], [O11]. It is closely related to the DMRG quantum computations [V03], [W93] and Hierarchical Tucker (HT) tensor decomposition [HK09], [H12].

We hope that our present paper will motivate further study of randomized augmentation and dual augmentation, their link to aggregation processes (cf. [MP80], [PQa]), specification of our methods to ill conditioned matrices that have displacement or rank structures (cf. [KKM79], [P90],

Table 10.4: The condition numbers $\kappa(A)$ of random circulant matrices A

| n | min | mean | max | std |
|---------|-------------------|-------------------|-------------------|-------------------|
| 256 | 9.6×10^0 | 1.1×10^2 | 3.5×10^3 | 4.0×10^2 |
| 512 | 1.4×10^1 | 8.5×10^1 | 1.1×10^3 | 1.3×10^2 |
| 1024 | 1.9×10^1 | 1.0×10^2 | 5.9×10^2 | 8.6×10^1 |
| 2048 | 4.2×10^1 | 1.4×10^2 | 5.7×10^2 | 1.0×10^2 |
| 4096 | 6.0×10^1 | 2.6×10^2 | 3.5×10^3 | 4.2×10^2 |
| 8192 | 9.5×10^1 | 3.0×10^2 | 1.5×10^3 | 2.5×10^2 |
| 16384 | 1.2×10^2 | 4.2×10^2 | 3.6×10^3 | 4.5×10^2 |
| 32768 | 2.3×10^2 | 7.5×10^2 | 5.6×10^3 | 7.1×10^2 |
| 65536 | 2.4×10^2 | 1.0×10^3 | 1.2×10^4 | 1.3×10^3 |
| 131072 | 3.9×10^2 | 1.4×10^3 | 5.5×10^3 | 9.0×10^2 |
| 262144 | 6.3×10^2 | 3.7×10^3 | 1.1×10^5 | 1.1×10^4 |
| 524288 | 8.0×10^2 | 3.2×10^3 | 3.1×10^4 | 3.7×10^3 |
| 1048576 | 1.2×10^3 | 4.8×10^3 | 3.1×10^4 | 5.1×10^3 |

[GKO95], [P01], [VBHK01], [EG99], [VVM07], [VVM08] on the displacement and rank structured matrices), and comparison and combination of our techniques with each other and various known methods, such as the techniques of [R90] (cf. our Remark 9.6), the homotopy continuation methods and Newton’s structured iteration (cf. Section 8.3, [PKRK], [P01, Chapter 6], and [P10]). Formal and experimental support of weakly randomized preconditioning (using fewer random parameters) remains an important research challenge. Natural extensions of our study should also include lower estimates for the smallest singular value of a random $m \times n$ matrix where $m \gg n$ or $n \gg m$, further links between augmentation and additive preprocessing, and the refinement, development and testing of our algorithms.

Table 10.5: Preconditioning tests

| Type | r | Cond (C) |
|------|-----|--------------|
| 1n | 1 | 3.21E+2 |
| 1n | 2 | 4.52E+3 |
| 1n | 4 | 2.09E+5 |
| 1n | 8 | 6.40E+2 |
| 1s | 1 | 5.86E+2x |
| 1s | 2 | 1.06E+4 |
| 1s | 4 | 1.72E+3 |
| 1s | 8 | 5.60E+3 |
| 2n | 1 | 8.05E+1 |
| 2n | 2 | 6.82E+3 |
| 2n | 4 | 2.78E+4 |
| 2n | 8 | 3.59E+3 |
| 2s | 1 | 1.19E+3 |
| 2s | 2 | 1.96E+3 |
| 2s | 4 | 1.09E+4 |
| 2s | 8 | 9.71E+3 |
| 3n | 1 | 2.02E+4 |
| 3n | 2 | 1.53E+3 |
| 3n | 4 | 6.06E+2 |
| 3n | 8 | 5.67E+2 |
| 3s | 1 | 2.39E+4 |
| 3s | 4 | 1.69E+3 |
| 3s | 8 | 6.74E+3 |
| 4n | 1 | 4.93E+2 |
| 4n | 2 | 4.48E+2 |
| 4n | 4 | 2.65E+2 |
| 4n | 8 | 1.64E+2 |
| 4s | 1 | 1.45E+3 |
| 4s | 2 | 5.11E+2 |
| 4s | 4 | 7.21E+2 |
| 4s | 8 | 2.99E+2 |

Table 10.6: Relative residual norms: randomized circulant GENP for well conditioned linear systems with ill conditioned leading blocks (cf. [PQZa, Table 2])

| dimension | iterations | min | max | mean | std |
|-----------|------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 64 | 0 | 4.7×10^{-14} | 8.0×10^{-11} | 4.0×10^{-12} | 1.1×10^{-11} |
| 64 | 1 | 1.9×10^{-15} | 5.3×10^{-13} | 2.3×10^{-14} | 5.4×10^{-14} |
| 256 | 0 | 1.7×10^{-12} | 1.4×10^{-7} | 2.0×10^{-9} | 1.5×10^{-8} |
| 256 | 1 | 8.3×10^{-15} | 4.3×10^{-10} | 4.5×10^{-12} | 4.3×10^{-11} |
| 1024 | 0 | 1.7×10^{-10} | 4.4×10^{-9} | 1.4×10^{-9} | 2.1×10^{-9} |
| 1024 | 1 | 3.4×10^{-14} | 9.9×10^{-14} | 6.8×10^{-14} | 2.7×10^{-14} |

Table 10.7: Tails of the SVDs and lower-rank approximations (cf. [PQ10])

| r | $\kappa(C)$ or rn_i | n | min | max | mean | std |
|-----|------------------------------|-----|------------------------|------------------------|------------------------|------------------------|
| 1 | $\kappa(C)$ | 64 | $2.38 \times 10^{+02}$ | $1.10 \times 10^{+05}$ | $6.25 \times 10^{+03}$ | $1.68 \times 10^{+04}$ |
| 1 | $\kappa(C)$ | 128 | $8.61 \times 10^{+02}$ | $7.48 \times 10^{+06}$ | $1.32 \times 10^{+05}$ | $7.98 \times 10^{+05}$ |
| 1 | $\kappa(C)$ | 256 | $9.70 \times 10^{+02}$ | $3.21 \times 10^{+07}$ | $3.58 \times 10^{+05}$ | $3.21 \times 10^{+06}$ |
| 1 | rn_1 | 64 | 4.01×10^{-10} | 1.50×10^{-07} | 5.30×10^{-09} | 1.59×10^{-08} |
| 1 | rn_1 | 128 | 7.71×10^{-10} | 5.73×10^{-07} | 1.58×10^{-08} | 6.18×10^{-08} |
| 1 | rn_1 | 256 | 7.57×10^{-10} | 3.2×10^{-07} | 1.69×10^{-08} | 5.02×10^{-08} |
| 1 | rn_2 | 64 | 4.01×10^{-10} | 1.50×10^{-07} | 5.30×10^{-09} | 1.59×10^{-08} |
| 1 | rn_2 | 128 | 7.71×10^{-10} | 5.73×10^{-07} | 1.58×10^{-08} | 6.18×10^{-08} |
| 1 | rn_2 | 256 | 7.57×10^{-10} | 3.22×10^{-07} | 1.69×10^{-08} | 5.02×10^{-08} |
| 8 | $\kappa(C)$ | 64 | $1.26 \times 10^{+03}$ | $1.61 \times 10^{+07}$ | $2.68 \times 10^{+05}$ | $1.71 \times 10^{+06}$ |
| 8 | $\kappa(C)$ | 128 | $2.92 \times 10^{+03}$ | $3.42 \times 10^{+06}$ | $1.58 \times 10^{+05}$ | $4.12 \times 10^{+05}$ |
| 8 | $\kappa(C)$ | 256 | $1.39 \times 10^{+04}$ | $8.75 \times 10^{+07}$ | $1.12 \times 10^{+06}$ | $8.74 \times 10^{+06}$ |
| 8 | rn_1 | 64 | 3.39×10^{-10} | 2.27×10^{-06} | 2.74×10^{-08} | 2.27×10^{-07} |
| 8 | rn_1 | 128 | 4.53×10^{-10} | 1.91×10^{-07} | 1.03×10^{-08} | 2.79×10^{-08} |
| 8 | rn_1 | 256 | 8.74×10^{-10} | 1.73×10^{-07} | 7.86×10^{-09} | 1.90×10^{-08} |
| 8 | rn_2 | 64 | 1.54×10^{-09} | 7.59×10^{-06} | 8.87×10^{-08} | 7.58×10^{-07} |
| 8 | rn_2 | 128 | 1.82×10^{-09} | 7.27×10^{-07} | 2.95×10^{-08} | 8.57×10^{-08} |
| 8 | rn_2 | 256 | 2.62×10^{-09} | 3.89×10^{-07} | 2.27×10^{-08} | 5.01×10^{-08} |
| 32 | $\kappa(C)$ | 64 | $1.77 \times 10^{+03}$ | $9.68 \times 10^{+06}$ | $1.58 \times 10^{+05}$ | $9.70 \times 10^{+05}$ |
| 32 | $\kappa(C)$ | 128 | $1.65 \times 10^{+04}$ | $6.12 \times 10^{+07}$ | $1.02 \times 10^{+06}$ | $6.19 \times 10^{+06}$ |
| 32 | $\kappa(C)$ | 256 | $3.57 \times 10^{+04}$ | $2.98 \times 10^{+08}$ | $4.12 \times 10^{+06}$ | $2.98 \times 10^{+07}$ |
| 32 | rn_1 | 64 | 2.73×10^{-10} | 3.29×10^{-08} | 2.95×10^{-09} | 4.93×10^{-09} |
| 32 | rn_1 | 128 | 3.94×10^{-10} | 1.29×10^{-07} | 7.18×10^{-09} | 1.64×10^{-08} |
| 32 | rn_1 | 256 | 6.80×10^{-10} | 4.00×10^{-07} | 1.16×10^{-08} | 4.27×10^{-08} |
| 32 | rn_2 | 64 | 2.10×10^{-09} | 1.49×10^{-07} | 1.55×10^{-08} | 2.18×10^{-08} |
| 32 | rn_2 | 128 | 2.79×10^{-09} | 3.80×10^{-07} | 3.81×10^{-08} | 6.57×10^{-08} |
| 32 | rn_2 | 256 | 5.35×10^{-09} | 1.05×10^{-06} | 5.70×10^{-08} | 1.35×10^{-07} |

Table 10.8: Heads of SVDs and low-rank approximations via dual additive preprocessing

| q | $\kappa(C_-)$ or rrn_i | n | min | max | mean | std |
|-----|---------------------------------|-----|------------------------|------------------------|------------------------|------------------------|
| 1 | $\kappa(C_-)$ | 64 | $1.83 \times 10^{+02}$ | $1.26 \times 10^{+06}$ | $1.74 \times 10^{+04}$ | $1.27 \times 10^{+05}$ |
| 1 | $\kappa(C_-)$ | 128 | $6.75 \times 10^{+02}$ | $8.76 \times 10^{+05}$ | $2.35 \times 10^{+04}$ | $9.10 \times 10^{+04}$ |
| 1 | $\kappa(C_-)$ | 256 | $4.19 \times 10^{+03}$ | $5.82 \times 10^{+05}$ | $4.43 \times 10^{+04}$ | $8.98 \times 10^{+04}$ |
| 1 | $\text{rn}^{(1)}$ | 64 | 2.43×10^{-10} | 3.86×10^{-08} | 2.55×10^{-09} | 5.43×10^{-09} |
| 1 | $\text{rn}^{(1)}$ | 128 | 4.36×10^{-10} | 1.15×10^{-07} | 4.45×10^{-09} | 1.24×10^{-08} |
| 1 | $\text{rn}^{(1)}$ | 256 | 6.40×10^{-10} | 3.17×10^{-08} | 4.00×10^{-09} | 5.16×10^{-09} |
| 1 | $\text{rn}^{(2)}$ | 64 | 8.30×10^{-10} | 3.86×10^{-08} | 2.81×10^{-09} | 5.35×10^{-09} |
| 1 | $\text{rn}^{(2)}$ | 128 | 1.21×10^{-9} | 1.15×10^{-07} | 4.80×10^{-09} | 1.23×10^{-08} |
| 1 | $\text{rn}^{(2)}$ | 256 | 1.72×10^{-9} | 3.18×10^{-08} | 4.53×10^{-09} | 4.97×10^{-09} |
| 8 | $\kappa(C_-)$ | 64 | $1.37 \times 10^{+03}$ | $1.87 \times 10^{+06}$ | $7.57 \times 10^{+04}$ | $2.16 \times 10^{+05}$ |
| 8 | $\kappa(C_-)$ | 128 | $3.80 \times 10^{+03}$ | $8.64 \times 10^{+06}$ | $2.00 \times 10^{+05}$ | $8.73 \times 10^{+05}$ |
| 8 | $\kappa(C_-)$ | 256 | $2.57 \times 10^{+04}$ | $1.54 \times 10^{+07}$ | $7.25 \times 10^{+05}$ | $2.03 \times 10^{+06}$ |
| 8 | $\text{rn}^{(1)}$ | 64 | 1.87×10^{-9} | 4.48×10^{-07} | 2.29×10^{-08} | 5.20×10^{-08} |
| 8 | $\text{rn}^{(1)}$ | 128 | 3.04×10^{-09} | 3.73×10^{-07} | 2.72×10^{-08} | 5.83×10^{-08} |
| 8 | $\text{rn}^{(1)}$ | 256 | 3.78×10^{-09} | 2.01×10^{-06} | 4.81×10^{-08} | 2.02×10^{-07} |
| 8 | $\text{rn}^{(2)}$ | 64 | 1.30×10^{-09} | 2.47×10^{-07} | 1.09×10^{-08} | 2.70×10^{-08} |
| 8 | $\text{rn}^{(2)}$ | 128 | 1.85×10^{-09} | 1.50×10^{-07} | 1.36×10^{-08} | 2.75×10^{-08} |
| 8 | $\text{rn}^{(2)}$ | 256 | 2.19×10^{-09} | 1.10×10^{-06} | 2.36×10^{-08} | 1.10×10^{-07} |
| 32 | $\kappa(C_-)$ | 64 | $3.75 \times 10^{+03}$ | $3.25 \times 10^{+07}$ | $6.01 \times 10^{+05}$ | $3.28 \times 10^{+06}$ |
| 32 | $\kappa(C_-)$ | 128 | $2.41 \times 10^{+04}$ | $1.09 \times 10^{+08}$ | $1.95 \times 10^{+06}$ | $1.10 \times 10^{+07}$ |
| 32 | $\kappa(C_-)$ | 256 | $1.33 \times 10^{+05}$ | $2.11 \times 10^{+10}$ | $2.18 \times 10^{+08}$ | $2.11 \times 10^{+09}$ |
| 32 | $\text{rn}^{(1)}$ | 64 | 7.78×10^{-09} | 1.39×10^{-06} | 8.17×10^{-08} | 1.94×10^{-07} |
| 32 | $\text{rn}^{(1)}$ | 128 | 9.81×10^{-09} | 2.35×10^{-06} | 1.17×10^{-07} | 3.05×10^{-07} |
| 32 | $\text{rn}^{(1)}$ | 256 | 2.05×10^{-08} | 3.99×10^{-06} | 1.91×10^{-07} | 5.06×10^{-07} |
| 32 | $\text{rn}^{(2)}$ | 64 | 1.84×10^{-09} | 2.62×10^{-07} | 1.85×10^{-08} | 4.09×10^{-08} |
| 32 | $\text{rn}^{(2)}$ | 128 | 2.47×10^{-09} | 6.77×10^{-07} | 2.93×10^{-08} | 8.38×10^{-08} |
| 32 | $\text{rn}^{(2)}$ | 256 | 5.05×10^{-09} | 8.85×10^{-07} | 4.38×10^{-08} | 1.14×10^{-07} |

Table 10.9: Heads of SVDs and low-rank approximation with random multipliers

| q | rrn_i | n | min | max | mean | std |
|-----|-------------------|-----|------------------------|------------------------|------------------------|------------------------|
| 1 | $\text{rn}^{(1)}$ | 64 | 2.35×10^{-10} | 1.32×10^{-07} | 3.58×10^{-09} | 1.37×10^{-08} |
| 1 | $\text{rn}^{(1)}$ | 128 | 4.41×10^{-10} | 3.28×10^{-08} | 3.55×10^{-09} | 5.71×10^{-09} |
| 1 | $\text{rn}^{(1)}$ | 256 | 6.98×10^{-10} | 5.57×10^{-08} | 5.47×10^{-09} | 8.63×10^{-09} |
| 1 | $\text{rn}^{(2)}$ | 64 | 8.28×10^{-10} | 1.32×10^{-07} | 3.86×10^{-09} | 1.36×10^{-08} |
| 1 | $\text{rn}^{(2)}$ | 128 | 1.21×10^{-09} | 3.28×10^{-08} | 3.91×10^{-09} | 5.57×10^{-09} |
| 1 | $\text{rn}^{(2)}$ | 256 | 1.74×10^{-09} | 5.58×10^{-08} | 5.96×10^{-09} | 8.47×10^{-09} |
| 8 | $\text{rn}^{(1)}$ | 128 | 2.56×10^{-09} | 1.16×10^{-06} | 4.30×10^{-08} | 1.45×10^{-07} |
| 8 | $\text{rn}^{(1)}$ | 256 | 4.45×10^{-09} | 3.32×10^{-07} | 3.40×10^{-08} | 5.11×10^{-08} |
| 8 | $\text{rn}^{(2)}$ | 64 | 1.46×10^{-09} | 9.56×10^{-08} | 5.77×10^{-09} | 1.06×10^{-08} |
| 8 | $\text{rn}^{(2)}$ | 128 | 1.64×10^{-09} | 4.32×10^{-07} | 1.86×10^{-08} | 5.97×10^{-08} |
| 8 | $\text{rn}^{(2)}$ | 256 | 2.50×10^{-09} | 1.56×10^{-07} | 1.59×10^{-08} | 2.47×10^{-08} |
| 32 | $\text{rn}^{(1)}$ | 64 | 6.80×10^{-09} | 2.83×10^{-06} | 1.01×10^{-07} | 3.73×10^{-07} |
| 32 | $\text{rn}^{(1)}$ | 128 | 1.25×10^{-08} | 6.77×10^{-06} | 1.28×10^{-07} | 6.76×10^{-07} |
| 32 | $\text{rn}^{(1)}$ | 256 | 1.85×10^{-08} | 1.12×10^{-06} | 1.02×10^{-07} | 1.54×10^{-07} |
| 32 | $\text{rn}^{(2)}$ | 64 | 1.84×10^{-09} | 6.50×10^{-07} | 2.30×10^{-08} | 8.28×10^{-08} |
| 32 | $\text{rn}^{(2)}$ | 128 | 3.11×10^{-09} | 1.45×10^{-06} | 2.87×10^{-08} | 1.45×10^{-07} |
| 32 | $\text{rn}^{(2)}$ | 256 | 4.39×10^{-09} | 2.16×10^{-07} | 2.37×10^{-08} | 3.34×10^{-08} |

Table 10.10: Heads of SVDs and low-rank approximations with random Toeplitz multipliers

| q | $\text{rrn}^{(i)}$ | n | min | max | mean | std |
|-----|--------------------|-----|------------------------|------------------------|------------------------|------------------------|
| 8 | $\text{rrn}^{(1)}$ | 64 | 2.22×10^{-09} | 7.89×10^{-06} | 1.43×10^{-07} | 9.17×10^{-07} |
| 8 | $\text{rrn}^{(1)}$ | 128 | 3.79×10^{-09} | 4.39×10^{-05} | 4.87×10^{-07} | 4.39×10^{-06} |
| 8 | $\text{rrn}^{(1)}$ | 256 | 5.33×10^{-09} | 3.06×10^{-06} | 6.65×10^{-08} | 3.12×10^{-07} |
| 8 | $\text{rrn}^{(2)}$ | 64 | 1.13×10^{-09} | 3.66×10^{-06} | 6.37×10^{-08} | 4.11×10^{-07} |
| 8 | $\text{rrn}^{(2)}$ | 128 | 1.81×10^{-09} | 1.67×10^{-05} | 1.90×10^{-07} | 1.67×10^{-06} |
| 8 | $\text{rrn}^{(2)}$ | 256 | 2.96×10^{-09} | 1.25×10^{-06} | 2.92×10^{-08} | 1.28×10^{-07} |
| 32 | $\text{rrn}^{(1)}$ | 64 | 6.22×10^{-09} | 5.00×10^{-07} | 4.06×10^{-08} | 6.04×10^{-08} |
| 32 | $\text{rrn}^{(1)}$ | 128 | 2.73×10^{-08} | 4.88×10^{-06} | 2.57×10^{-07} | 8.16×10^{-07} |
| 32 | $\text{rrn}^{(1)}$ | 256 | 1.78×10^{-08} | 1.25×10^{-06} | 1.18×10^{-07} | 2.03×10^{-07} |
| 32 | $\text{rrn}^{(2)}$ | 64 | 1.64×10^{-09} | 1.26×10^{-07} | 9.66×10^{-09} | 1.48×10^{-08} |
| 32 | $\text{rrn}^{(2)}$ | 128 | 5.71×10^{-09} | 9.90×10^{-07} | 5.50×10^{-08} | 1.68×10^{-07} |
| 32 | $\text{rrn}^{(2)}$ | 256 | 4.02×10^{-09} | 2.85×10^{-07} | 2.74×10^{-08} | 4.48×10^{-08} |

Table 10.11: Relative residual norms: ill conditioned linear systems via nmb approximation and block triangulation

| n | \mathbf{r} | min | max | mean | std |
|-----|--------------|------------------------|-----------------------|------------------------|------------------------|
| 32 | 1 | 1.49×10^{-13} | 1.36×10^{-9} | 4.25×10^{-11} | 1.56×10^{-10} |
| 32 | 2 | 3.70×10^{-13} | 2.13×10^{-8} | 3.83×10^{-10} | 2.35×10^{-9} |
| 32 | 4 | 9.33×10^{-13} | 1.08×10^{-8} | 3.37×10^{-10} | 1.26×10^{-9} |
| 64 | 1 | 1.11×10^{-12} | 6.87×10^{-9} | 2.03×10^{-10} | 7.49×10^{-10} |
| 64 | 2 | 1.53×10^{-12} | 1.21×10^{-8} | 5.86×10^{-10} | 1.77×10^{-9} |
| 64 | 4 | 2.21×10^{-12} | 1.27×10^{-7} | 1.69×10^{-9} | 1.28×10^{-8} |

Table 10.12: Relative residual norms: ill conditioned linear systems with MLDIVIDE(A,B)

| n | r | min | max | mean | std |
|-----|-----|-----------------------|--------------------|-----------------------|--------------------|
| 32 | 1 | 6.34×10^{-3} | 7.44×10^1 | 1.74×10^0 | 7.53×10^0 |
| 32 | 2 | 2.03×10^{-2} | 1.32×10^1 | 9.19×10^{-1} | 1.62×10^0 |
| 32 | 4 | 4.57×10^{-2} | 1.36×10^1 | 1.14×10^0 | 1.93×10^0 |
| 64 | 1 | 3.82×10^{-3} | 9.93×10^0 | 1.03×10^0 | 1.66×10^0 |
| 64 | 2 | 1.96×10^{-2} | 1.27×10^2 | 3.09×10^0 | 1.40×10^1 |
| 64 | 4 | 7.13×10^{-3} | 6.63×10^0 | 8.23×10^{-1} | 1.20×10^0 |

Table 10.13: Relative residual norms: ill conditioned linear systems with dual additive preprocessing and block triangulation

| n | q | min | max | mean | std |
|-----|-----|------------------------|------------------------|------------------------|------------------------|
| 32 | 1 | 2.33×10^{-14} | 2.28×10^{-06} | 2.31×10^{-08} | 2.28×10^{-07} |
| 32 | 2 | 3.40×10^{-13} | 4.93×10^{-08} | 9.11×10^{-10} | 5.71×10^{-09} |
| 32 | 4 | 5.97×10^{-13} | 1.63×10^{-07} | 2.22×10^{-09} | 1.64×10^{-08} |
| 64 | 1 | 3.90×10^{-14} | 2.78×10^{-05} | 2.81×10^{-07} | 2.78×10^{-06} |
| 64 | 2 | 3.53×10^{-13} | 3.76×10^{-08} | 1.13×10^{-09} | 4.72×10^{-09} |
| 64 | 4 | 3.54×10^{-12} | 2.53×10^{-07} | 5.19×10^{-09} | 2.83×10^{-08} |

Table 10.14: Relative residual norms: ill conditioned linear system with random multipliers and block triangulation

| n | q | min | max | mean | std |
|-----|-----|------------------------|------------------------|------------------------|------------------------|
| 32 | 1 | 7.08×10^{-30} | 4.00×10^{-23} | 4.52×10^{-25} | 4.01×10^{-24} |
| 32 | 2 | 7.49×10^{-30} | 2.29×10^{-21} | 2.77×10^{-23} | 2.33×10^{-22} |
| 32 | 4 | 1.46×10^{-28} | 1.63×10^{-07} | 4.83×10^{-25} | 2.73×10^{-24} |
| 64 | 1 | 1.13×10^{-29} | 1.01×10^{-24} | 2.31×10^{-26} | 1.11×10^{-25} |
| 64 | 2 | 6.60×10^{-29} | 6.90×10^{-24} | 1.45×10^{-25} | 7.73×10^{-25} |
| 64 | 4 | 2.60×10^{-28} | 1.41×10^{-21} | 1.61×10^{-23} | 1.42×10^{-22} |

Table 10.15: Relative residual norms: ill conditioned linear systems with the SMW formula

| n | r | min | max | mean | std |
|-----|-----|------------------------|------------------------|------------------------|------------------------|
| 64 | 1 | 1.18×10^{-15} | 6.30×10^{-13} | 2.37×10^{-14} | 7.45×10^{-14} |
| 64 | 2 | 3.42×10^{-15} | 1.94×10^{-10} | 2.15×10^{-12} | 1.94×10^{-11} |
| 64 | 4 | 6.66×10^{-15} | 1.25×10^{-10} | 1.82×10^{-12} | 1.25×10^{-11} |
| 128 | 1 | 5.79×10^{-15} | 4.85×10^{-12} | 1.21×10^{-13} | 4.96×10^{-13} |
| 128 | 2 | 1.45×10^{-14} | 1.85×10^{-11} | 5.23×10^{-13} | 1.88×10^{-12} |
| 128 | 4 | 8.41×10^{-14} | 4.75×10^{-11} | 2.89×10^{-12} | 5.95×10^{-12} |

Table 10.16: The CPU time (in cycles) for solving an ill conditioned real symmetric Toeplitz linear system

| n | Alg. 6.1 | QR | SVD | QR/Alg. 6.1 | SVD/Alg. 6.1 |
|------|----------|---------|---------|-------------|--------------|
| 512 | 56.3 | 148.4 | 4134.8 | 2.6 | 73.5 |
| 1024 | 120.6 | 1533.5 | 70293.1 | 12.7 | 582.7 |
| 2048 | 265.0 | 11728.1 | — | 44.3 | — |
| 4096 | 589.4 | — | — | — | — |
| 8192 | 1304.8 | — | — | — | — |

Appendix

A Operations with structured matrices in terms of their displacements

The following simple theorem can be found in [P00] or [P01, Section 1.5].

Theorem A.1. *Assume five matrices A, B, C, M and N and a pair of scalars α and β . Then as long as the matrix sizes are compatible we have*

$$A(\alpha M + \beta N) - (\alpha M + \beta N)B = \alpha(AM - MB) + \beta(AN - NB), \quad (\text{A.1})$$

$$A^T M^T - B^T M^T = -(BM - MA)^T, \quad (\text{A.2})$$

$$A(MN) - (MN)C = (AM - MB)N + M(BN - NC). \quad (\text{A.3})$$

Furthermore for a nonsingular matrix M we have

$$AM^{-1} - M^{-1}B = -M^{-1}(BM - MA)M^{-1}. \quad (\text{A.4})$$

Corollary A.1. *Under the assumptions of Theorem A.1 we have*

$$G_{A,B}(\alpha M + \beta N) = (\alpha G_{A,B}(M) \mid \beta G_{A,B}(N)), \quad (\text{A.5})$$

$$H_{A,B}(\alpha M + \beta N) = (\alpha H_{A,B}(M) \mid \beta H_{A,B}(N)), \quad (\text{A.6})$$

$$G_{A,B}(M^T) = -H_{B^T, A^T}(M^T), \quad H_{A,B}(M^T) = G_{B^T, A^T}(M^T), \quad (\text{A.7})$$

$$G_{A,C}(MN) = (G_{A,B}(M) \mid MG_{B,C}(N)), \quad (\text{A.8})$$

$$H_{A,C}(MN) = (N^T H_{A,B}(M) \mid H_{B,C}(N)), \quad (\text{A.9})$$

$$G_{A,B}(M^{-1}) = -M^{-1}G_{B,A}(M), \quad H_{A,B}(M^{-1}) = M^{-T}H_{B,A}(M) \quad (\text{A.10})$$

and consequently

$$d_{A,B}(\alpha M + \beta N) \leq d_{A,B}(M) + d_{A,B}(N), \quad (\text{A.11})$$

$$d_{A,B}(M^T) = d_{B^T, A^T}(M), \quad (\text{A.12})$$

$$d_{A,C}(MN) \leq d_{A,B}(M) + d_{B,C}(N), \quad (\text{A.13})$$

$$d_{A,B}(M^{-1}) = d_{B,A}(M). \quad (\text{A.14})$$

B Uniform random sampling and nonsingularity of random matrices

Uniform random sampling of elements from a finite set Δ is their selection from this set at random, independently of each other and under the uniform probability distribution on the set Δ .

Theorem B.1. *Under the assumptions of Lemma 3.1 let the values of the variables of the polynomial be randomly and uniformly sampled from a finite set Δ . Then the polynomial vanishes with a probability at most $\frac{d}{|\Delta|}$.*

Corollary B.1. *Let the entries of a general or Toeplitz $m \times n$ matrix have been randomly and uniformly sampled from a finite set Δ of cardinality $|\Delta|$ (in any fixed ring). Let $l = \min\{m, n\}$. Then (a) every $k \times k$ submatrix M for $k \leq l$ is nonsingular with a probability at least $1 - \frac{k}{|\Delta|}$ and (b) is strongly nonsingular with a probability at least $1 - \sum_{i=1}^k \frac{i}{|\Delta|} = 1 - \frac{(k+1)k}{2|\Delta|}$. Furthermore (c) if the submatrix M is indeed nonsingular, then any entry of its inverse is nonzero with a probability at least $1 - \frac{k-1}{|\Delta|}$.*

Proof. The claimed properties of nonsingularity and nonvanishing hold for generic matrices. The singularity of a $k \times k$ matrix means that its determinant vanishes, but the determinant is a polynomial of total degree k in the entries. Therefore Theorem B.1 implies parts (a) and consequently (b). Part (c) follows because a fixed entry of the inverse vanishes if and only if the respective entry of the adjoint vanishes, but up to the sign the latter entry is the determinant of a $(k-1) \times (k-1)$ submatrix of the input matrix M , and so it is a polynomial of degree $k-1$ in its entries. \square

C Extremal singular values of random complex matrices

We have assumed dealing with real random matrices and vectors throughout the paper, but most of our study can be readily extended to the computations in the field \mathbb{C} of complex numbers if we replace the transposes A^T by the Hermitian transposes A^H . All the results of Section 3.2 apply to this case equally well. Below we elaborate upon the respective extension of our probabilistic bounds on the norms and singular values.

Definition C.1. *The set $\mathcal{G}_{\mathbb{C}, \mu, \sigma}^{m \times n}$ of complex Gaussian random $m \times n$ matrices with a mean μ and a variance σ is the set $\{A + B\sqrt{-1}\}$ for $(A \mid B) \in \mathcal{G}_{\mu, \sigma}^{m \times 2n}$ (cf. Definition 3.2).*

We can immediately extend Theorem 3.2 to the latter matrices. Let us extend Theorem 3.1. Its original proof in [SST06] relies on the following result.

Lemma C.1. *Suppose y is a positive number; $\mathbf{w} \in \mathbb{R}^{n \times 1}$ is any fixed real unit vector, $\|\mathbf{w}\| = 1$, $A \in \mathcal{G}_{\mu, \sigma}^{n \times n}$ and therefore is nonsingular with probability 1. Then*

$$\text{Probability}\{\|A^{-1}\mathbf{w}\| > 1/y\} \leq \sqrt{\frac{2}{\pi}} \frac{y}{\sigma} \text{ for } j = 1, \dots, n.$$

The following lemma and corollary extend Lemmas 3.2 and C.1 to the complex case.

Lemma C.2. *The bound of Lemma 3.2 also holds provided $\mathbf{t} = \mathbf{q} + \mathbf{r}\sqrt{-1}$ is a fixed complex unit vector and $\mathbf{b} = \mathbf{f} + \mathbf{g}\sqrt{-1} \in \mathcal{G}_{\mathbb{C}, \mu, \sigma}^{n \times 1}$ is a complex vector such that \mathbf{f} , \mathbf{g} , \mathbf{q} and \mathbf{r} are real vectors, $\|\mathbf{t}\| = 1$, and the vectors \mathbf{f} and \mathbf{g} are in $\mathcal{G}_{\mu, \sigma}^{n \times 1}$.*

Proof. We have $\mathbf{t}^H \mathbf{b} = \mathbf{q}^T \mathbf{f} + \mathbf{r}^T \mathbf{g} + (\mathbf{q}^T \mathbf{g} - \mathbf{r}^T \mathbf{f})\sqrt{-1}$, and so $|\mathbf{t}^H \mathbf{b}|^2 = |\mathbf{q}^T \mathbf{f} + \mathbf{r}^T \mathbf{g}|^2 + |\mathbf{q}^T \mathbf{g} - \mathbf{r}^T \mathbf{f}|^2$. Hence $|\mathbf{t}^H \mathbf{b}| \geq |\mathbf{q}^T \mathbf{f} + \mathbf{r}^T \mathbf{g}| = |\mathbf{u}^T \mathbf{v}|$ where $\mathbf{u}^T = (\mathbf{q}^T \mid \mathbf{r}^T)$ and $\mathbf{v} = (\mathbf{f}^T \mid \mathbf{g}^T)^T$. Note that $\mathbf{v} \in \mathcal{G}_{\mu, \sigma}^{1 \times 2n}$ and $\|\mathbf{u}\| = \|\mathbf{t}\| = 1$ and apply Lemma 3.2 to real vectors \mathbf{u} and \mathbf{v} replacing \mathbf{b} and \mathbf{t} . \square

Corollary C.1. *Suppose y is a positive number and suppose a matrix $A \in \mathcal{G}_{\mathbb{C}, \mu, \sigma}^{n \times n}$ and therefore is nonsingular with probability 1. Then*

$$\text{Probability}\{\|A^{-1}\mathbf{e}_j\| > 1/y\} \leq \sqrt{\frac{2}{\pi}} \frac{y}{\sigma} \text{ for } j = 1, \dots, n.$$

Proof. In the case of real matrices A the corollary is supported by the argument in the proof of [SST06, Lemma 3.2], which employs the well known estimate that we state as our Lemma 3.2. Now we employ Lemma C.2 instead of this estimate, otherwise keep the same argument as in [SST06], and arrive at Corollary C.1. \square

Corollary C.2. *Under the assumptions of Corollary C.1 we have $\|A^{-1}\| \leq \sum_{j=1}^n X_j$ where X_j are nonnegative random variables such that*

$$\text{Probability}\{X_j > 1/y\} \leq \sqrt{\frac{2}{\pi}} \frac{y}{\sigma} \text{ for } j = 1, \dots, n.$$

Proof. Recall that for any $n \times n$ matrix B we have $\|B\| = \|B\mathbf{w}\|$ for some unit vector $\mathbf{w} = \sum_{j=1}^n w_j \mathbf{e}_j$. We have $|w_j| \leq \|\mathbf{w}\| = 1$ for all j . Substitute $B = A^{-1}$ and obtain $\|A^{-1}\| = \|A^{-1}\mathbf{w}\| = \|\sum_{j=1}^n A^{-1} w_j \mathbf{e}_j\| \leq \sum_{j=1}^n |w_j| \|A^{-1}\mathbf{e}_j\|$, and so $\|A^{-1}\| \leq \sum_{j=1}^n X_j$ where $X_j = \|A^{-1}\mathbf{e}_j\|$ for all j . It remains to combine this bound with Corollary C.1. \square

The corollary implies that $\text{Probability}\{|A^{-1}| > 1/y\}$ converges to 0 proportionally to y as $y \rightarrow 0$, which can be viewed as an extension of Theorem 3.1 to the case of complex inputs. One can deduce similar extensions of Theorems 3.4–3.6. The resulting estimates are a little weaker than in Section 3.3, being overly pessimistic; actually random complex matrices are a little better conditioned than random real matrices (see [E88], [ES05], [CD05] and our Table 10.2).

Acknowledgement: Our research has been supported by NSF Grant CCF-1116736 and PSC CUNY Awards 63153–0041 and 64512–0042. We are also very grateful to Professor Dr. Siegfried M. Rump and Mr. Jesse Wolf for helpful comments.

References

- [A94] O. Axelsson, *Iterative Solution Methods*, Cambridge Univ. Press, England, 1994.
- [B80] D. Bini, Border Rank of $p \times q \times 2$ Tensors and the Optimal Approximation of a Pair of Bilinear Forms, in *Lecture Notes in Computer Science*, **85**, 98–108, Springer, 1980.
- [B85] D. Bini, Tensor and Border Rank of Certain Classes of Matrices and the Fast Evaluation of Determinant, Inverse Matrix and Eigenvalues, *Calcolo*, **22**, 209–228, 1985.
- [B86] D. Bini, Border Rank of $m \times n \times (mn - q)$ Tensors, *Linear Algebra and Its Applications*, **79**, 45–51, 1986.
- [B02] M. Benzi, Preconditioning Techniques for Large Linear Systems: a Survey, *J. of Computational Physics*, **182**, 418–477, 2002.
- [B11] C. Beltrán, Estimates on the Condition Number of Random, Rank-deficient Matrices, *IMA Journal of Numerical Analysis*, **31**, **1**, 25–39, 2011.
- [BC87] D. Bini, M. Capovani, Tensor Rank and Border Rank of Band Toeplitz Matrices, *SIAM J. on Computing*, **2**, 252–258, 1987.
- [BCLR79] D. Bini, M. Capovani, G. Lotti, F. Romani, $O(n^{2.7799})$ Complexity for $n \times n$ Approximate Matrix Multiplication, *Information Processing Letters*, **8**, 234–235, 1979.
- [BG05] A. Böttcher, S. M. Grudsky, *Spectral Properties of Banded Toeplitz Matrices*, SIAM Publications, Philadelphia, 2005.
- [BP94] D. Bini, V. Y. Pan, *Polynomial and Matrix Computations*, volume 1: Fundamental Algorithms, Birkhäuser, Boston, 1994.
- [CD05] Z. Chen, J. J. Dongarra, Condition Numbers of Gaussian Random Matrices, *SIAM J. on Matrix Analysis and Applications*, **27**, 603–620, 2005.
- [CDG03] B. Carpentieri, I. S. Duff, L. Giraud, A Class of Spectral Two-level Preconditioners, *SIAM J. Scientific Computing*, **25**, **2**, 749–765, 2003.
- [CPW74] R. E. Cline, R. J. Plemmons, and G. Worm, Generalized Inverses of Certain Toeplitz Matrices, *Linear Algebra and Its Applications*, **8**, 25–33, 1974.
- [CW90] Coppersmith, S. Winograd, Matrix Multiplication via Arithmetic Progressions. *J. Symbolic Comput.*, **9**, **3**, 251–280, 1990.
- [D83] J. D. Dixon, Estimating Extremal Eigenvalues and Condition Numbers of Matrices, *SIAM J. on Numerical Analysis*, **20**, **4**, 812–814, 1983.
- [D88] J. Demmel, The Probability That a Numerical Analysis Problem Is Difficult, *Math. of Computation*, **50**, 449–480, 1988.
- [DL78] R. A. Demillo, R. J. Lipton, A Probabilistic Remark on Algebraic Program Testing, *Information Processing Letters*, **7**, **4**, 193–195, 1978.

- [DS01] K. R. Davidson, S. J. Szarek, Local Operator Theory, Random Matrices, and Banach Spaces, in *Handbook on the Geometry of Banach Spaces* (W. B. Johnson and J. Lindenstrauss editors), pages 317–368, North Holland, Amsterdam, 2001.
- [E88] A. Edelman, Eigenvalues and Condition Numbers of Random Matrices, *SIAM J. on Matrix Analysis and Applications*, **9**, 4, 543–560, 1988.
- [EG99] Y. Eidelman, I. Gohberg, On a New Class of Structured Matrices, *Integral Equations and Operator Theory*, **34**, 293–324, Birkhäuser, Basel, 1999.
- [ES05] A. Edelman, B. D. Sutton, Tails of Condition Number Distributions, *SIAM J. on Matrix Analysis and Applications*, **27**, 2, 547–560, 2005.
- [G97] A. Greenbaum, *Iterative Methods for Solving Linear Systems*, SIAM, Philadelphia, 1997.
- [GE96] M. Gu, S. C. Eisenstat, Efficient algorithms for computing a strong rank-revealing QR factorization, *SIAM Journal on Scientific Computing*, **17**, 848–869, 1996.
- [GK72] I. Gohberg, N. Y. Krupnick, A Formula for the Inversion of Finite Toeplitz Matrices, *Matematicheskie Issledovaniia* (in Russian), **7**, 2, 272–283, 1972.
- [GKO95] I. Gohberg, T. Kailath, V. Olshevsky, Fast Gaussian Elimination with Partial Pivoting for Matrices with Displacement Structure, *Math. of Comp.*, **64**(212), 1557–1576, 1995.
- [GL96] G. H. Golub, C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, Baltimore, Maryland, 1996 (third addition).
- [GOS08] S. A. Goreinov, I. V. Oseledets, D. V. Savostyanov, E. E. Tyrtyshnikov, N. L. Zamarashkin, How to Find a Good Submatrix, Research Report 08-10, ICM HKBU, Kowloon Tong, Hong Kong, 2008.
- [GS72] I. Gohberg, A. Sementsul, On the Inversion of Finite Toeplitz Matrices and Their Continuous Analogs, *Matematicheskie Issledovaniia* (in Russian), **7**, 2, 187–224, 1972.
- [GT01] S. A. Goreinov, E. E. Tyrtyshnikov, The Maximal-volume Concept in Approximation by Low-rank Matrices, *Contemporary Mathematics*, **208**, 47–51, 2001.
- [GTZ97] S. A. Goreinov, E. E. Tyrtyshnikov, N. L. Zamarashkin, A Theory of Pseudo-skeleton Approximations, *Linear Algebra and Its Applications*, **261**, 1–22, 1997.
- [H02] N. J. Higham, *Accuracy and Stability in Numerical Analysis*, SIAM, Philadelphia, 2002 (second edition).
- [H12] W. Hackbusch, *Tensor spaces and numerical tensor calculus*, Springer, Berlin, 2012.
- [HK09] W. Hackbusch, S. Khn, A New Scheme for the Tensor Representation, *J. Fourier Anal. Appls.*, **15**, 5, 706–722 2009.
- [HMT11] N. Halko, P. G. Martinsson, J. A. Tropp, Finding Structure with Randomness: Probabilistic Algorithms for Constructing Approximate Matrix Decompositions, *SIAM Review*, **53**, 2, 217–288, 2011.
- [HP92] Y. P. Hong, C.-T. Pan, The rank revealing QR decomposition and SVD, *Math. of Computation*, **58**, 213–232, 1992.
- [K04] I. Kaporin, The Aggregation and Cancellation Techniques as a Practical Tool for Faster Matrix Multiplication, *Theoretical Computer Science*, **315**, 2–3, 469–510, 2004.
- [KB09] T. G. Kolda, B. W. Bader, Tensor Decompositions and Applications, *SIAM Review*, **51**, 3, 455–500, 2009.

- [KKM79] T. Kailath, S. Y. Kung, M. Morf, Displacement Ranks of Matrices and Linear Equations, *Journal Math. Analysis and Appls*, **68**, **2**, 395–407, 1979.
- [KL94] C. S. Kenney, A. J. Laub, Small-Sample Statistical Condition Estimates for General Matrix Functions, *SIAM J. on Scientific and Statistical Computing*, **15**, 36–61, 1994.
- [KV99] P. Kravanja, M. Van Barel, Algorithms for Solving Rational Interpolation Problems Related to Fast and Superfast Solvers for Toeplitz Systems, *SPIE*, 359–370, 1999.
- [LPS92] J. Laderman, V. Y. Pan, H. X. Sha, On Practical Algorithms for Accelerated Matrix Multiplication, *Linear Algebra and Its Applications*, **162–164**, 557–588, 1992.
- [MMD08] M. W. Mahoney, M. Maggioni, P. Drineas, Tensor-CUR Decompositions for Tensor-based Data, *SIAM J. on Matrix Analysis and Applications*, **30**, **2**, 957–987, 2008.
- [MP80] W. L. Miranker, V. Y. Pan, Methods of Aggregations, *Linear Algebra and Its Applications*, **29**, 231–257, 1980.
- [ORO05] T. Ogita, S. M. Rump, S. Oishi, Accurate Sum and Dot Product, *SIAM Journal on Scientific Computing*, **26**, **6**, 1955–1988, 2005.
- [O09] I. V. Oseledets, A Compact Matrix Form of the d-Dimensional Tensor Decomposition, Preprint 2009-01, *INM RAS*, March 2009.
- [O11] I. V. Oseledets, Tensor-Train Decomposition, *SIAM J. Scientific Comp.*, **33**, **5**, 2295–2317, 2011.
- [OT09] I. V. Oseledets, E. E. Tyrtyshnikov, Breaking the Curse of Dimensionality, or How to Use SVD in Many Dimensions, *SIAM J. Scientific Comp.*, **31**, **5**, 3744–3759, 2009.
- [OT10] I. V. Oseledets, E. E. Tyrtyshnikov, TT-cross Approximation for Multidimensional Arrays, *Linear Algebra Appls*. **432**, **1**, 70–88, 2010.
- [OT11] I. Oseledets, E. E. Tyrtyshnikov, Algebraic Wavelet Transform via Quantics Tensor Train Decomposition, *SIAM J. Sci. Comp.*, **33**, **3**, 1315–1328, 2011.
- [P72] V. Y. Pan, On Schemes for the Evaluation of Products and Inverses of Matrices (in Russian), *Uspekhi Matematicheskikh Nauk*, **27**, **5 (167)**, 249–250, 1972.
- [P79] V. Y. Pan, Fields Extension and Trilinear Aggregating, Uniting and Canceling for the Acceleration of Matrix Multiplication, *Proceedings of the 20th Annual IEEE Symposium on Foundations of Computer Science (FOCS 79)*, 28–38, IEEE Computer Society Press, Long Beach, California, 1979.
- [P81] V. Y. Pan, New Combinations of Methods for the Acceleration of Matrix Multiplications, *Computers and Mathematics (with Applications)*, **7**, 73–125, 1981.
- [P84] V. Y. Pan, How Can We Speed up Matrix Multiplication? *SIAM Review*, **26**, **3**, 393–415, 1984.
- [P90] V. Y. Pan, On Computations with Dense Structured Matrices, *Math. of Computation*, **55**, **191**, 179–190, 1990. Proceedings version in *Proc. Intern. Symposium on Symbolic and Algebraic Computation (ISSAC’89)*, 34–42, ACM Press, New York, 1989.
- [P92] V. Y. Pan, Parallel Solution of Toeplitz-like Linear Systems, *J. of Complexity*, **8**, 1–21, 1992.
- [P93] V. Y. Pan, Concurrent Iterative Algorithm for Toeplitz-like Linear Systems, *IEEE Transactions on Parallel and Distributed Systems*, **4**, **5**, 592–600, 1993.

- [P93a] V. Y. Pan, Decreasing the Displacement Rank of a Matrix, *SIAM Journal on Matrix Analysis and Applications*, **14**, **1**, 118–121, 1993.
- [P00] V. Y. Pan, Nearly Optimal Computations with Structured Matrices, *Proceedings of 11th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'2000)*, 953–962, ACM Press, New York, and SIAM Publications, Philadelphia, 2000.
- [P00a] C.-T. Pan, On the Existence and Computation of Rank-revealing LU Factorization, *Linear Algebra and Its Applications*, **316**, 199–222, 2000.
- [P01] V. Y. Pan, *Structured Matrices and Polynomials: Unified Superfast Algorithms*, Birkhäuser/Springer, Boston/New York, 2001.
- [P07] V. Y. Pan, New Homotopic/Factorization and Symmetrization Techniques for Newton's and Newton/Structured Iteration, *Computer and Math. with Applications*, **54**, 721–729, 2007.
- [P10] V. Y. Pan, Newton's Iteration for Matrix Inversion, Advances and Extensions, pp. 364–381, in *Matrix Methods: Theory, Algorithms and Applications* (dedicated to the Memory of Gene Golub, edited by V. Olshevsky and E. Tyrtyshnikov), World Scientific Publishing, New Jersey, ISBN-13 978-981-283-601-4, ISBN-10-981-283-601-2 (2010).
- [P11] V.Y. Pan, Nearly optimal solution of rational linear systems of equations with symbolic lifting and numerical initialization. *Computers & Math. with Applications*, **62**, 1685–1706, 2011.
- [PBRZ99] V. Y. Pan, S. Brahnman, B. Murphy, R. E. Rosholt, Newton's Iteration for Structured Matrices and Linear Systems of Equations, *SIAM Volume on Fast Reliable Algorithms for Matrices with Structure* (T. Kailath and A. H. Sayed, editors), 189–210, SIAM Publications, Philadelphia, 1999.
- [PGMQ] V. Y. Pan, D. Grady, B. Murphy, G. Qian, R. E. Rosholt, A. Ruslanov, Schur Aggregation for Linear Systems and Determinants, *Theoretical Computer Science, Special Issue on Symbolic-Numerical Algorithms* (D. A. Bini, V. Y. Pan, and J. Verschelde editors), **409**, **2**, 255–268, 2008.
- [PKRK] V. Y. Pan, M. Kunin, R. Rosholt, H. Kodal, Homotopic Residual Correction Processes, *Math. of Computation*, **75**, 345–368, 2006.
- [PIMR10] V. Y. Pan, D. Ivolgin, B. Murphy, R. E. Rosholt, Y. Tang, X. Yan, Additive Preconditioning for Matrix Computations, *Linear Algebra and Its Applications*, **432**, 1070–1089, 2010.
- [PQ10] V. Y. Pan, G. Qian, Randomized Preprocessing of Homogeneous Linear Systems of Equations, *Linear Algebra and Its Applications*, **432**, 3272–3318, 2010.
- [PQa] V. Y. Pan, G. Qian, Solving Linear Systems of Equations with Randomization, Augmentation and Aggregation II, Tech. Report TR 2012007, *Ph.D. Program in Computer Science, Graduate Center, the City University of New York*, 2012.
Available at <http://www.cs.gc.cuny.edu/tr/techreport.php?id=434>
- [PQZa] V. Y. Pan, G. Qian, A. Zheng, Randomized Preprocessing versus Pivoting, *Linear Algebra and Its Applications*, in print. <http://dx.doi.org/10.1016/j.laa.2011.02.052>
- [PQZC] V. Y. Pan, G. Qian, A. Zheng, Z. Chen, Matrix Computations and Polynomial Root-finding with Preprocessing, *Linear Algebra and Its Applications*, **434**, 854–879, 2011.
- [PRW02] V. Y. Pan, Y. Rami, X. Wang, Structured Matrices and Newtons Iteration: Unified Approach, *Linear Algebra and Its Applications*, **343/344**, 233–265, 2002.

- [PS91] V. Y. Pan, R. Schreiber, An Improved Newton Iteration for the Generalized Inverse of a Matrix, with Applications, *SIAM Journal on Scientific and Statistical Computing*, **12**, 5, 1109–1131, 1991.
- [R90] S. M. Rump, Approximate Inverses of Almost Singular Matrices Still Contain Useful Information, Technical Report 90.1, *Forschungsschwerpunkt Informations-und Kommunikationstechnik*, TU Hamburg-Harburg, 1990.
- [S80] J. T. Schwartz, Fast Probabilistic Algorithms for Verification of Polynomial Identities, *Journal of ACM*, **27**, 4, 701–717, 1980.
- [S98] G. W. Stewart, *Matrix Algorithms, Vol I: Basic Decompositions*, SIAM, 1998.
- [S10] A. J. Stothers, On the Complexity of Matrix Multiplication, Ph.D. Thesis, University of Edinburgh, 2010.
- [SST06] A. Sankar, D. Spielman, S.-H. Teng, Smoothed Analysis of the Condition Numbers and Growth Factors of Matrices, *SIAM J. on Matrix Analysis*, **28**, 2, 446–476, 2006.
- [T00] E. E. Tyrtysnikov, Incomplete Cross Approximation in Mosaic Skeleton Method, *Computing*, **64**, 367–380, 2000.
- [V99] M. Van Barel, A Superfast Toeplitz Solver, 1999.
Available at <http://www.cs.kuleuven.be/~marc/software/index.html>
- [V03] G. Vidal, Efficient classical simulation of slightly entangled quantum computations, *Phys. Rev. Letters*, **91** (14), 147902-1 147902-4, 2003.
- [VBHK01] M. Van Barel, G. Heinig, P. Kravanja, A Stabilized Superfast Solver for Nonsymmetric Toeplitz Systems, *SIAM J. on Matrix Analysis and Applications*, **23**, 2, 494–510, 2001.
- [VK98] M. Van Barel, P. Kravanja, A Stabilized Superfast Solver for Indefinite Hankel Systems, *Linear Algebra and its Applications*, **284**, 1–3, 335–355, 1998.
- [VVM07] R. Vandebril, M. Van Barel, N. Mastronardi, *Matrix Computations and Semiseparable Matrices: Linear Systems* (Volume 1), The Johns Hopkins University Press, Baltimore, Maryland, 2007.
- [VVM08] R. Vandebril, M. Van Barel, N. Mastronardi, *Matrix Computations and Semiseparable Matrices: Eigenvalue and Singular Value Methods* (Volume 2), The Johns Hopkins University Press, Baltimore, Maryland, 2008.
- [VW12] V. Vassilevska Williams, Multiplying Matrices Faster than Coppersmith–Winograd, *Proc. 44th Annual ACM Symposium on Theory of Computing (STOC 2012)*, 887–898, ACM Press, New York, 2012.
- [W93] S.R. White, Density-matrix algorithms for quantum renormalization groups, *Phys. Rev. B*, **48** (14), 10345–10356, 1993.
- [W07] X. Wang, Affect of Small Rank Modification on the Condition Number of a Matrix, *Computer and Math. (with Applications)*, **54**, 819–825, 2007.
- [Z79] R. E. Zippel, Probabilistic Algorithms for Sparse Polynomials, *Proceedings of EURO-SAM’79, Lecture Notes in Computer Science*, **72**, 216–226, Springer, Berlin, 1979.